Efficient reduced-basis treatment of nonaffine and nonlinear partial differential equations

Martin A. Grepl, Yvon Maday, N.C. Nguyen, Anthony T. Patera

Mathematical Modelling and Numerical Analysis Will be set by the publisher
Modélisation Mathématique et Analyse Numérique

EFFICIENT REDUCED-BASIS TREATMENT OF NONAFFINE AND NONLINEAR PARTIAL DIFFERENTIAL EQUATIONS

M.A. Grepl[1], Y. Maday[2], N.C. Nguyen[3] and A.T. Patera[4]

Abstract. In this paper, we extend the reduced-basis approximations developed earlier for *linear* elliptic and parabolic partial differential equations with *affine* parameter dependence to problems involving (a) *nonaffine* dependence on the parameter, and (b) *nonlinear* dependence on the field variable. The method replaces the nonaffine and nonlinear terms with a coefficient function approximation which then permits an efficient offline-online computational decomposition. We first review the coefficient function approximation procedure: the essential ingredients are (i) a good collateral reduced-basis approximation space, and (ii) a stable and inexpensive interpolation procedure. We then apply this approach to linear nonaffine and nonlinear elliptic and parabolic equations; in each instance, we discuss the reduced-basis approximation and the associated offline-online computational procedures. Numerical results are presented to assess our approach.

1991 Mathematics Subject Classification. 35J25,35J60,35K15,35K55.

March 29, 2006.

1. Introduction

The design, optimization, control, and characterization of engineering components or systems often requires repeated, reliable, and real-time prediction of selected performance metrics, or "outputs," s^{e} [1]; typical

Keywords and phrases: Reduced-basis methods, parametrized PDEs, non-affine parameter dependence, offine-online procedures, elliptic PDEs, parabolic PDEs, nonlinear PDEs

[1] Massachusetts Institute of Technology, Room 3-264, Cambridge, MA USA

[2] Université Pierre et Marie Curie-Paris6, UMR 7598 Laboratoire Jacques-Louis Lions, B.C. 187, Paris, F-75005 France, E-mail: maday@ann.jussieu.fr; and Division of Applied Mathematics, Brown University

[3] Massachusetts Institute of Technology, Room 37-435, Cambridge, MA USA

[4] Massachusetts Institute of Technology, Room 3-266, Cambridge, MA USA

[1]Here superscript "e" shall refer to "exact." We shall later introduce a "truth approximation" which will bear no superscript.

"outputs" include forces, critical stresses or strains, flowrates, or heat fluxes. These outputs are typically functionals of a field variable, $u^{\rm e}(\mu)$ — such as temperatures or velocities — associated with a parametrized partial differential equation that describes the underlying physics; the parameters, or "inputs," μ, serve to identify a particular configuration of the component or system — geometry, material properties, boundary conditions, and loads. The relevant system behavior is thus described by an implicit input-output relationship, $s^{\rm e}(\mu)$, evaluation of which demands solution of the underlying partial differential equation (PDE).

The abstract formulation for an elliptic problem can be stated as follows: given any $\mu \in \mathcal{D} \subset \mathbb{R}^P$, we evaluate $s^{\rm e}(\mu) = \ell(u^{\rm e}(\mu))$, where $u^{\rm e}(\mu) \in X^{\rm e}$ is the solution of

$$a(u^{\rm e}(\mu), v; \mu) = f(v; \mu), \quad \forall v \in X^{\rm e}. \tag{1}$$

Here $\mathcal{D}$ is the parameter domain in which our P-tuple (input) parameter μ resides; $X^{\rm e}(\Omega)$ is an appropriate Hilbert space; Ω is a bounded domain in $\mathbb{R}^d$ with Lipschitz continuous boundary $\partial\Omega$; $f(\cdot;\mu)$, $\ell(\cdot)$ are $X^{\rm e}$-continuous linear functionals; and $a(\cdot,\cdot;\mu)$ is a $X^{\rm e}$-continuous bilinear form.

In actual practice, of course, we do not have access to the exact solution; we thus replace $u^{\rm e}(\mu)$ with a "truth" approximation, $u(\mu)$, which resides in (say) a suitably fine piecewise-linear finite element approximation space $X \subset X^{\rm e}$ of *very* large dimension $\mathcal{N}$. Our "truth" approximation is thus: given any $\mu \in \mathcal{D}$, we evaluate $s(\mu) = \ell(u(\mu))$, where $u(\mu) \in X$ is the solution of

$$a(u(\mu), v; \mu) = f(v; \mu), \quad \forall v \in X. \tag{2}$$

We shall assume — hence the appellation "truth" — that the discretization is sufficiently rich such that $u(\mu)$ and $u^{\rm e}(\mu)$ and hence $s(\mu)$ and $s^{\rm e}(\mu)$ are indistinguishable at the accuracy level of interest. The reduced-basis approximation shall be built upon this reference (or "truth") finite element approximation, and the reduced-basis error will thus be evaluated with respect to $u(\mu) \in X$. Our formulation must be stable and efficient as $\mathcal{N} \to \infty$.

We now turn to the abstract formulation for the controlled parabolic case. For simplicity, in this paper we will directly consider a time-discrete framework associated to the time interval $I \equiv]0, t_f]$. We divide $\bar{I} \equiv [0, t_f]$ into K subintervals of equal length $\Delta t = \frac{t_f}{K}$ and define $t^k \equiv k\Delta t$, $0 \le k \le K \equiv \frac{t_f}{\Delta t}$, and $\mathbb{I} \equiv \{t^0, \ldots, t^K\}$;

for notational convenience, we also introduce $\mathbb{K} \equiv \{1, \ldots, K\}$. We shall consider Euler-Backward for the time integration; we can also readily treat higher-order schemes such as Crank-Nicolson [12]. The "truth" approximation is thus: given any $\mu \in \mathcal{D}$, we evaluate the output $s(\mu, t^k) = \ell(u(\mu, t^k)),\ \forall k \in \mathbb{K}$, where $u(\mu, t^k) \in X$ satisfies

$$m(u(\mu, t^k), v) + \Delta t\, a(u(\mu, t^k), v; \mu) = m(u(\mu, t^{k-1}), v) + \Delta t\, f(v; \mu)\, b(t^k), \quad \forall v \in X,\ \forall k \in \mathbb{K}, \tag{3}$$

with initial condition (say) $u(\mu, t^0) = u_0(\mu) = 0$. Here, $f(\cdot, \mu)$ and $\ell(\cdot)$ are Y^{e}-continuous ($X^{\mathrm{e}} \subset Y^{\mathrm{e}}$) linear functionals, $m(\cdot, \cdot)$ is a Y^{e}-continuous bilinear form, and $b(t^k)$ is the control input. We note that the output, $s(\mu, t^k)$, and the field variable, $u(\mu, t^k)$, are now functions of the discrete time t^k, $\forall k \in \mathbb{K}$.

Our goal is the development of numerical methods that permit the *rapid* yet *accurate and reliable* prediction of these PDE-induced input-output relationships in *real-time* or *in the limit of many queries* — relevant, for example, in the design, optimization, control, and characterization contexts. To achieve this goal we will pursue the reduced-basis method. The reduced-basis method was first introduced in the late 1970s for the nonlinear analysis of structures [1, 25] and subsequently abstracted and analyzed [5, 11, 28, 33] and extended [16, 18, 26] to a much larger class of parametrized partial differential equations. The foundation of the reduced basis method is the realization that, in many instances, the set of all solutions $u(\mu)$ (say, in the elliptic case) as μ varies can be approximated very well by its projection on a finite and low dimensional vector space: for sufficiently well chosen μ_i, there exist coefficients $c_i = c_i^N(\mu)$ such that the finite sum $\sum_{i=1}^N c_i u(\mu_i)$ is very close to $u(\mu)$ for any μ.

More recently, the reduced-basis approach and also associated *a posteriori* error estimation procedures have been successfully developed for (i) linear elliptic and parabolic PDEs that are affine in the parameter [13, 20, 21, 29, 40] — the bilinear form $a(w, v; \mu)$ can be expressed as

$$a(w, v; \mu) = \sum_{q=1}^{Q} \Theta^q(\mu)\, a^q(w, v), \tag{4}$$

where the $\Theta^q : \mathcal{D} \to \mathbb{R}$ and $a^q(w, v), 1 \le q \le Q$, are *parameter dependent* functions and *parameter-independent* bilinear forms, respectively; and (ii) elliptic PDEs that are at most quadratically nonlinear in the first argument [24, 38, 39] — in particular, $a(w, v; \mu)$ satisfies (4) and is at most quadratic in w (but

of course linear in v). In these cases a very efficient offline-online computational strategy relevant in the many-query and real-time contexts can be developed. The operation count for the online stage — in which, given a new parameter value, we calculate the reduced-basis output and associated error bound — depends on a low power of the dimension of the reduced-basis space N (typically small) and Q; but it is *independent* of $\mathcal{N}$, the dimension of the underlying "truth" finite element approximation.

Unfortunately, if a is not affine in the parameter this computational strategy breaks down; the online complexity will still depend on $\mathcal{N}$. For example, for *general* $g(x;\mu)$ (here $x \in \Omega$ and $\mu \in \mathcal{D}$), the bilinear form

$$a(w,v;\mu) \equiv \int_\Omega \nabla w \cdot \nabla v + \int_\Omega g(x;\mu)\, w\, v \tag{5}$$

will not admit an efficient (online $\mathcal{N}$-independent) computational decomposition. In a recent CRAS note [4], we introduce a technique that recovers the efficient offline-online decomposition even in the presence of nonaffine parameter dependence. In this approach, we develop a "collateral" reduced-basis expansion $g_M(x;\mu)$ for $g(x;\mu)$ and then replace $g(x;\mu)$ in (5) with some necessarily affine approximation $g_M(x;\mu) = \sum_{m=1}^{M} \varphi_{M\,m}(\mu) q_m(x)$. The essential ingredients are (*i*) a "good" collateral reduced-basis approximation space, $W_M^g = \text{span}\{q_m(x), 1 \le m \le M\}$, (*ii*) a stable and inexpensive ($\mathcal{N}$-independent) interpolation procedure by which to determine the $\varphi_{M\,m}(\mu), 1 \le m \le M$, and (*iii*) an effective *a posteriori* estimator with which to quantify the newly introduced error terms. In this paper we shall expand upon the brief presentation in [4] and furthermore address the treatment of nonaffine *parabolic* problems; we shall also extend the technique to elliptic and parabolic problems in which g is a nonaffine *nonlinear* function of the field variable u — we hence treat certain classes of nonlinear problems.

A large number of model order reduction (MOR) techniques [2, 7, 8, 22, 27, 32, 36, 41] have been developed to treat nonlinear time-dependent problems. One approach is linearization [41] and polynomial approximation [8, 27]. However, inefficient representation of the nonlinear terms and fast exponential growth (with the degree of the nonlinear approximation order) of the computational complexity render these methods quite expensive, in particular for strong nonlinearities; other approaches for highly nonlinear systems (such as piecewise-linearization) [32, 35] suffer from similar drawbacks. It is also important to note that most

MOR techniques focus *only* on temporal variations; the development of reduced-order models for parametric applications — our focus here — is much less common [6, 9].

This paper is organized as follows: In Section 2 we present a short review of the "empirical interpolation method" – coefficient function approximation — introduced in [4]. The abstract problem formulation, reduced-basis approximation, and computational considerations for linear coercive elliptic and linear coercive parabolic problems with nonaffine parameter dependence are then discussed in Section 3 and Section 4, respectively. We extend these results in Section 5 to monotonic nonlinear elliptic PDEs and and in Section 6 to monotonic nonlinear parabolic PDEs. Numerical results are included in each section in order to confirm and assess our theoretical results. (Note that, due to space limitations, we do not present in this paper associated *a posteriori* error estimators; the reader is referred to [4, 12, 23, 37] for a detailed development of this topic.)

2. Empirical Interpolation

2.1. Coefficient–Function Procedure

We begin by summarizing the results in [4]. We consider the problem of approximating a given μ-dependent function $g(\,\cdot\,;\mu) \in L^\infty(\Omega),\ \forall \mu \in \mathcal{D}$, of sufficient regularity by a reduced-basis expansion $g_M(\,\cdot\,;\mu)$; here, $L^\infty(\Omega) \equiv \{v \,|\, \text{ess sup}_{v\in\Omega}\, |v(x)| < \infty\}$. To this end, we introduce the nested sample sets $S_M^g = \{\mu_1^g \in \mathcal{D}, \ldots, \mu_M^g \in \mathcal{D}\}$, and associated nested reduced-basis spaces $W_M^g = \text{span}\,\{\xi_m \equiv g(x;\mu_m^g), 1 \le m \le M\}$, in which our approximation g_M shall reside. We also introduce the best approximation $g_M^*(\cdot;\mu) \equiv \arg\min_{z\in W_M^g} \|g(\cdot;\mu) - z\|_{L^\infty(\Omega)}$ and the associated error $\varepsilon_M^*(\mu) \equiv \|g(\cdot;\mu) - g_M^*(\cdot;\mu)\|_{L^\infty(\Omega)}$. (More generally, we can work in a Banach space B that in our context will be $L^\infty(\Omega)$ or $L^2(\Omega)$. Then $g \in C^0(\mathcal{D}; B)$ and the forthcoming construction of S_M^g is effected with respect to the B norm.)

The construction of S_M^g and W_M^g is based on a greedy selection process. To begin, we choose our first sample point to be $\mu_1^g = \arg\max_{\mu\in\Xi^g} \|g(\,\cdot\,;\mu)\|_{L^\infty(\Omega)}$, and define $S_1^g = \{\mu_1^g\}$, $\xi_1 \equiv g(x;\mu_1^g)$, and $W_1^g = \text{span}\{\xi_1\}$; here Ξ^g is a suitably large but finite-dimensional parameter set in $\mathcal{D}$. Then, for $M \ge 2$, we set $\mu_M^g = \arg\max_{\mu\in\Xi^g} \varepsilon_{M-1}^*(\mu)$, and define $S_M^g = S_{M-1}^g \cup \mu_M^g$, $\xi_M = g(x;\mu_M^g)$, and $W_M^g = \text{span}\,\{\xi_m,\, 1 \le m \le M\}$. In essence, W_M^g comprises basis functions from the parametrically induced manifold $\mathcal{M}^g \equiv \{g(\,\cdot\,;\mu) \,|\, \mu \in \mathcal{D}\}$.

Thanks to our truth approximation, the optimization for $g^*_{M-1}(\,\cdot\,;\mu)$ and hence $\varepsilon^*_{M-1}(\mu)$ is a *standard linear program*.

We note that the determination of μ^g_M requires the solution of a linear program for *each* parameter point in Ξ^g; the computational cost thus depends strongly on the size of Ξ^g. In the parabolic case this cost may be prohibitively large — at least in our current implementation — if the function g is time-varying either through an explicit dependence on time or (for nonlinear problems) an implicit dependence via the field variable $u(\mu, t^k)$. As we shall see, in these cases the parameter sample Ξ^g is in effect replaced by the *parameter-time* sample $\tilde{\Xi}^g \equiv \Xi^g \times \mathbb{I}$; even for modest K the computational cost can be very high. We thus propose an alternative construction of S^g_M: we replace the $L^\infty(\Omega)$-norm in our best approximation by the $L^2(\Omega)$-norm, where $L^2(\Omega)$ is the space of square integrable functions over Ω; our next sample point is now given by $\mu^g_M = \arg\max_{\mu\in\Xi^g} \inf_{z\in W^g_{M-1}} \|g(\,\cdot\,;\mu) - z\|_{L^2(\Omega)}$, which is relatively inexpensive to evaluate — the computational cost to evaluate $\inf_{z\in W^g_{M-1}} \|g(\,\cdot\,;\mu) - z\|_{L^2(\Omega)}$ is $O(M\mathcal{N}) + O(M^3)$. The following analysis is still rigorous for this alternative (or "surrogate") construction of S^g_M, since we are working in a finite-dimensional space and hence all norms are equivalent; in fact, the $L^\infty(\Omega)$ and $L^2(\Omega)$ procedures yield very similar convergence results in practice (see Section 2.3).

We begin the analysis of our greedy procedure with the following Lemma.

Lemma 2.1. *Suppose that $M_{\max}$ is chosen such that the dimension of span $\mathcal{M}^g$ exceeds $M_{\max}$; then, for any $M \le M_{\max}$, the space W^g_M is of dimension M.*

Proof. It directly follows from our hypothesis on $M_{\max}$ that $\varepsilon_0 \equiv \varepsilon^*_{M_{\max}}(\mu^g_{M_{\max}+1}) > 0$; our "arg max" construction then implies $\varepsilon^*_{M-1}(\mu^g_M) \ge \varepsilon_0$, $2 \le M \le M_{\max}$, since $\varepsilon^*_{M-1}(\mu^g_M) \ge \varepsilon^*_{M-1}(\mu^g_{M+1}) \ge \varepsilon^*_M(\mu^g_{M+1})$. We now prove lemma 2.1 by induction. Clearly, $\dim(W^g_1) = 1$; assume $\dim(W^g_{M-1}) = M-1$; then if $\dim(W^g_M) \ne M$, we have $g(\,\cdot\,;\mu^g_M) \in W^g_{M-1}$ and thus $\varepsilon^*_{M-1}(\mu^g_M) = 0$; however, the latter contradicts $\varepsilon^*_{M-1}(\mu^g_M) \ge \varepsilon_0 > 0$. □

We now construct nested sets of interpolation points $T_M = \{x_1, \ldots, x_M\}$, $1 \le M \le M_{\max}$. We first set $x_1 = \arg\operatorname{ess\,sup}_{x\in\Omega} |\xi_1(x)|$, $q_1 = \xi_1(x)/\xi_1(x_1)$, $B^1_{11} = 1$. Then for $M = 2, \ldots, M_{\max}$, we solve the linear system $\sum_{j=1}^{M-1} \sigma^{M-1}_j q_j(x_i) = \xi_M(x_i)$, $1 \le i \le M-1$, and set $r_M(x) = \xi_M(x) - \sum_{j=1}^{M-1} \sigma^{M-1}_j q_j(x)$,

$x_M = \arg\,\text{ess sup}_{x\in\Omega} |r_M(x)|$, $q_M(x) = r_M(x)/r_M(x_M)$, and $B^M_{i\,j} = q_j(x_i)$, $1 \le i, j \le M$. It remains to demonstrate

Lemma 2.2. *The construction of the interpolation points is well-defined, and the functions* $\{q_1, \ldots, q_M\}$ *form a basis for* W^g_M. *In addition, the matrix* B^M *is lower triangular with unity diagonal.*

Proof. We shall proceed by induction. Clearly, we have $W^g_1 = \text{span}\,\{q_1\}$. Next we assume $W^g_{M-1} = \text{span}\,\{q_1, \ldots, q_{M-1}\}$; if (i) B^{M-1} is invertible and (ii) $|r_M(x_M))| > 0$, then our construction may proceed and we may form $W^g_M = \text{span}\,\{q_1, \ldots, q_M\}$. To prove (i), we just note from the construction procedure that $B^{M-1}_{i\,j} = r_j(x_i)/r_j(x_j) = 0$ for $i < j$; that $B^{M-1}_{i\,j} = r_j(x_i)/r_j(x_j) = 1$ for $i = j$; and that $\left|B^{M-1}_{i\,j}\right| = |r_j(x_i)/r_j(x_j)| \le 1$ for $i > j$ since $x_j = \arg\,\text{ess sup}_{x\in\Omega} |r_j(x)|, 1 \le j \le M$. Hence, B^M is lower triangular with unity diagonal. To prove (ii) (and hence also that the $x_i, 1 \le i \le M$, are distinct), we observe that $|r_M(x_M)| \ge \varepsilon^*_{M-1}(\mu^g_M) \ge \varepsilon_0 > 0$ since $\varepsilon^*_{M-1}(\mu^g_M)$ is the error associated with the best approximation. □

Furthermore, from the invertibility of B^M, we immediately derive

Lemma 2.3. *For any* M*-tuple* $(\alpha_i)_{i=1,\ldots,M}$ *of real numbers, there exists a unique element* $w \in W^g_M$ *such that* $w(x_i) = \alpha_i,\ 1 \le i \le M$.

It remains to develop an *efficient* procedure for obtaining a *good* collateral reduced-basis expansion $g_M(\cdot;\mu)$. Based on the approximation space W^g_M and set of interpolation points T_M, we can readily construct an approximation to $g(x;\mu)$. Indeed, our coefficient function approximation is the interpolant of g over T_M as provided for from Lemma 2.3:

$$g_M(x;\mu) = \sum_{m=1}^{M} \varphi_{M\,m}(\mu)\, q_m(x), \tag{6}$$

where $\varphi_M(\mu) \in \mathbb{R}^M$ is given by

$$\sum_{j=1}^{M} B^M_{i\,j}\, \varphi_{M\,j}(\mu) = g(x_i;\mu),\ 1 \le i \le M; \tag{7}$$

note that $g_M(x_i;\mu) = g(x_i;\mu), 1 \le i \le M$. We define the associated error as

$$\varepsilon_M(\mu) \equiv \|g(\,\cdot\,;\mu) - g_M(\,\cdot\,;\mu)\|_{L^\infty(\Omega)}. \tag{8}$$

It remains to understand how well $g_M(x;\mu)$ approximates $g(x;\mu)$.

2.2. **Error Analysis**

2.2.1. *A Priori Stability: Lebesgue Constant*

To begin, we define a "Lebesgue constant" [10, 30, 34] $\Lambda_M = \sup_{x\in\Omega} \sum_{m=1}^{M} |V_m^M(x)|$. Here, the $V_m^M(x) \in W_M^g$ are characteristic functions satisfying $V_m^M(x_n) = \delta_{mn}, 1 \le m, n \le M$, the existence and uniqueness of which is guaranteed by Lemma 2.3; here δ_{mn} is the Kronecker delta symbol. It can be shown that

Lemma 2.4. *The set of all characteristic functions* $\left\{V_m^M\right\}_{m=1}^M$ *is a basis for* W_M^g. *Furthermore, the two bases* $q_m, 1 \le m \le M$, *and* $V_m^M, 1 \le m \le M$, *are related by*

$$q_i(x) = \sum_{j=1}^{M} B_{j\,i}^M \, V_j^M(x), \quad 1 \le i \le M \ . \tag{9}$$

Proof. It is immediate from the definition of the V_m^M that the set of all characteristic functions $\left\{V_m^M\right\}_{m=1}^M$ is linearly independent. This set thus constitutes a basis for W_M^g, in fact a nodal basis associated with the set $\{x_m\}_{m=1}^M$. Then, we consider $x = x_n,\ 1 \le n \le M$, and note that $\sum_{j=1}^M \ B_{j\,i}^M \, V_j^M(x_n) = \sum_{j=1}^M \ B_{j\,i}^M \, \delta_{jn} = B_{n\,i}^M = q_i(x_n),\ 1 \le i \le M$; it thus follows from Lemma 2.3 that (9) holds. □

We observe that Λ_M depends on W_M^g and T_M, but not on μ. We can further prove

Lemma 2.5. *The interpolation error* $\varepsilon_M(\mu)$ *satisfies* $\varepsilon_M(\mu) \le \varepsilon_M^*(\mu)(1 + \Lambda_M), \ \forall\, \mu \in \mathcal{D}$.

Proof. We first introduce $e_M^*(x;\mu) = g(x;\mu) - g_M^*(x;\mu)$. It then follows that

$$\begin{aligned} g_M(x;\mu) - g_M^*(x;\mu) &= \sum_{m=1}^{M} \left(g_M(x_i;\mu) - g_M^*(x_i;\mu)\right) \, V_m^M(x) \\ &= \sum_{m=1}^{M} \left(\left(g_M(x_i;\mu) - g(x_i;\mu)\right) + \left(g(x_i;\mu) - g_M^*(x_i;\mu)\right)\right) V_m^M(x) \\ &= \sum_{m=1}^{M} e_M^*(x_i;\mu) \, V_m^M(x) \ . \end{aligned} \tag{10}$$

Furthermore, from the definition of $\varepsilon_M(\mu)$ and $\varepsilon_M^*(\mu)$, and the triangle inequality, we obtain

$$\varepsilon_M(\mu) = \|g(\,\cdot\,;\mu) - g_M(\,\cdot\,;\mu)\|_{L^\infty(\Omega)} \le \varepsilon_M^*(\mu) + \|g_M(\,\cdot\,;\mu) - g_M^*(\,\cdot\,;\mu)\|_{L^\infty(\Omega)}.$$

This yields, from (10),

$$\begin{aligned}\varepsilon_M(\mu) - \varepsilon_M^*(\mu) &\leq \|g_M(\,\cdot\,;\mu) - g_M^*(\,\cdot\,;\mu)\|_{L^\infty(\Omega)} \\ &= \|\sum_{i=1}^{M} e_M^*(x_i;\mu)\, V_i^M(x)\|_{L^\infty(\Omega)} \\ &\leq \max_{i\in\{1,\ldots,M\}} |e_M^*(x_i;\mu)|\, \Lambda_M;\end{aligned}$$

the desired result then immediately follows from $|e_M^*(x_i;\mu)| \leq \varepsilon_M^*(\mu)$, $1 \leq i \leq M$. □

We can further show

Proposition 2.6. *The Lebesgue constant Λ_M satisfies $\Lambda_M \leq 2^M - 1$.*

Proof. We first recall two crucial properties of the matrix B^M: (i) B^M is lower triangular with unity diagonal — $q_m(x_m) = 1$, $1 \leq m \leq M$, and (ii) all entries of B^M are of modulus no greater than unity — $\|q_m\|_{L^\infty(\Omega)} \leq 1$, $1 \leq m \leq M$. Hence, from (9) we can write

$$\begin{aligned}|V_m^M(x)| &= \left| q_m(x) - \sum_{i=m+1}^{M} B_{i\,m}^M V_i^M(x) \right| \\ &\leq 1 + \sum_{i=m+1}^{M} |V_i^M(x)|, \quad 1 \leq m \leq M-1.\end{aligned}$$

It follows that, starting from $|V_M^M(x)| = |q_M(x)| \leq 1$, we can deduce $|V_{M+1-m}^M(x)| \leq 1 + |V_M^M(x)| + \ldots + |V_{M+2-m}^M(x)| \leq 2^{m-1}$, $2 \leq m \leq M$, and thus obtain $\sum_{m=1}^{M} |V_m^M(x)| \leq 2^M - 1$. □

Proposition 2.6 is very pessimistic and of little practical value (though $\varepsilon_M^*(\mu)$ does often converge sufficiently rapidly that $\varepsilon_M^*(\mu)\, 2^M \to 0$ as $M \to \infty$); this is not surprising given analogous results in the theory of polynomial interpolation [10,30,34]. In applications, the actual asymptotic behavior of Λ_M is *much* lower than the upper bound of Proposition 2.6; however, Proposition 2.6 does provide a theoretical basis for some stability.

2.2.2. A Posteriori *Estimators*

Given an approximation $g_M(x;\mu)$ for $M \leq M_{\max} - 1$, we define $\mathcal{E}_M(x;\mu) \equiv \hat{\varepsilon}_M(\mu)\, q_{M+1}(x)$, where $\hat{\varepsilon}_M(\mu) \equiv |g(x_{M+1};\mu) - g_M(x_{M+1};\mu)|$. In general, $\varepsilon_M(\mu) \geq \hat{\varepsilon}_M(\mu)$, since $\varepsilon_M(\mu) = ||g(\cdot;\mu) - g_M(\cdot;\mu)||_{L^\infty(\Omega)} \geq |g(x;\mu) - g_M(x;\mu)|$ for all $x \in \Omega$, and thus also for $x = x_{M+1}$. However, we can prove

Proposition 2.7. *If* $g(\,\cdot\,;\mu) \in W^g_{M+1}$, *then (i)* $g(x;\mu) - g_M(x;\mu) = \pm\mathcal{E}_M(x;\mu)$, *and (ii)* $\|g(\,\cdot\,;\mu) - g_M(\,\cdot\,;\mu)\|_{L^\infty(\Omega)} = \hat{\varepsilon}_M(\mu)$.[2]

Proof. By our assumption $g(\,\cdot\,;\mu) \in W^g_{M+1}$, there exists $\kappa(\mu) \in \mathbb{R}^{M+1}$ such that $g(x;\mu) - g_M(x;\mu) = \sum_{m=1}^{M+1} \kappa_m(\mu)\, q_m(x)$. We now consider $x = x_i, 1 \le i \le M+1$, and arrive at

$$\sum_{m=1}^{M+1} \kappa_m(\mu)\, q_m(x_i) = g(x_i;\mu) - g_M(x_i;\mu), \quad 1 \le i \le M+1 \ .$$

It thus follows that $\kappa_m(\mu) = 0$, $1 \le m \le M$, since $g(x_i;\mu) - g_M(x_i;\mu) = 0, 1 \le i \le M$, and the matrix $q_m(x_i)(= B^M_{im})$ is lower triangular, and that $\kappa_{M+1}(\mu) = g(x_{M+1};\mu) - g_M(x_{M+1};\mu)$ since $q_{M+1}(x_{M+1}) = 1$; this concludes the proof of (i). The proof of (ii) then directly follows from $\|q_{M+1}\|_{L^\infty(\Omega)} = 1$. □

Of course, in general $g(\,\cdot\,;\mu) \notin W^g_{M+1}$, and hence our estimator $\hat{\varepsilon}_M(\mu)$ is unfortunately a lower bound. However, if $\varepsilon_M(\mu) \to 0$ very fast, we expect that the effectivity,

$$\eta_M(\mu) \equiv \frac{\hat{\varepsilon}_M(\mu)}{\varepsilon_M(\mu)} \quad , \tag{11}$$

shall be close to unity; furthermore, the estimator is very inexpensive – *one additional evaluation* of $g(\,\cdot\,;\mu)$ at a single point in Ω. (Note we can readily improve the rigor of our bound at only modest additional cost: if we assume that $g(;\mu) \in W^g_{M+k}$, then $\hat{\varepsilon}_M = 2^{k-1} \max_{i\in\{1,\ldots,k\}} |g(x_{M+k};\mu) - g_M(x_{M+k};\mu))|$ is an upper bound for $\varepsilon_M(\mu)$ (see Propositions 2.6 and 2.7).)

We refer to [4, 12, 23] for the incorporation of these error estimators into output bounds for reduced basis approximations of nonaffine partial differential equations.

2.3. **Numerical Results**

We consider the function $g(\cdot;\mu) = G(\cdot;\mu)$, where

$$G(x;\mu) \equiv \frac{1}{\sqrt{(x_{(1)} - \mu_{(1)})^2 + (x_{(2)} - \mu_{(2)})^2}} \tag{12}$$

for $x = (x_{(1)}, x_{(2)}) \in \Omega \equiv]0,1[^{\,2}$ and $\mu \in \mathcal{D} \equiv [-1, -0.01]^2$. We choose for Ξ^g a deterministic grid of 40×40 parameter points over $\mathcal{D}$. We take $\mu^g_1 = (-0.01, -0.01)$ and then pursue the empirical interpolation

[2]Note that the proof of (ii) $\|g(\,\cdot\,;\mu) - g_M(\,\cdot\,;\mu)\|_{L^\infty(\Omega)} \le \hat{\varepsilon}_M(\mu)$ in [4] is technically correct but in fact misleading since equality indeed always holds.

procedure described in Section 2.1 to construct S_M^g, W_M^g, T_M, and B^M, $1 \le M \le M_{\max}$, for $M_{\max} = 51$. We observe that the parameter points in S_M^g, shown in Figure 1(a), are mainly distributed around the corner $(-0.01, -0.01)$ of the parameter domain; and that the interpolation points in T_M, plotted in Figure 1(b), are largely allocated around the corner $(0, 0)$ of the physical domain Ω.

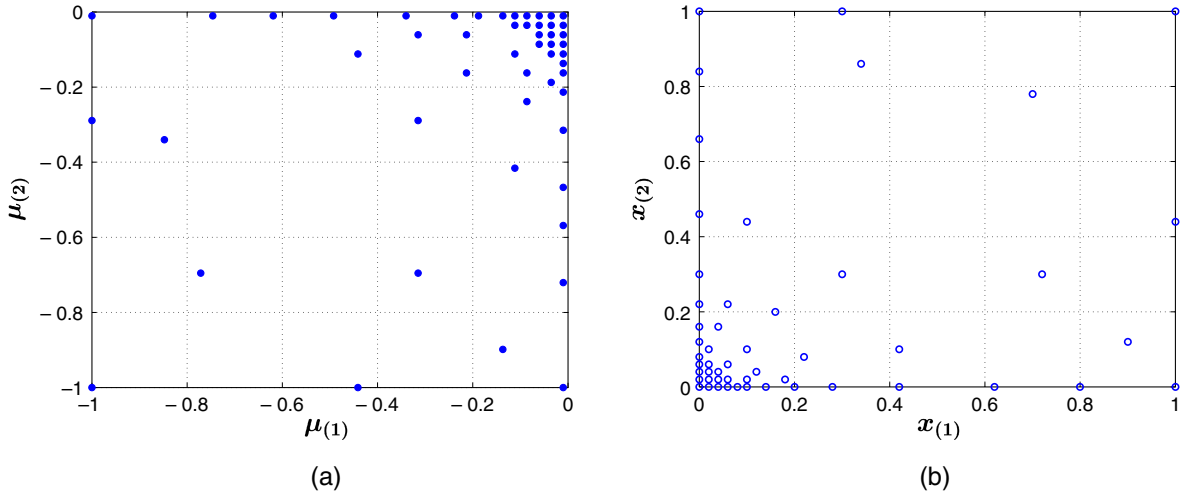


FIGURE 1. (a) Parameter sample set S_M^g, $M_{\max} = 51$, and (b) interpolation points x_m, $1 \le m \le M_{\max}$, for the function $G(x;\mu)$ of (12).

We now introduce a parameter test sample Ξ_{Test}^g of size $Q_{\text{Test}} = 225$, and define $\varepsilon_{M,\max} = \max_{\mu \in \Xi_{\text{Test}}^g} \varepsilon_M(\mu)$, $\varepsilon_{M,\max}^* = \max_{\mu \in \Xi_{\text{Test}}^g} \varepsilon_M^*(\mu)$, $\overline{\rho}_M = Q_{\text{Test}}^{-1} \sum_{\mu \in \Xi_{\text{Test}}^g} \left(\varepsilon_M(\mu) / (\varepsilon_M^*(\mu)(1 + \Lambda_M)) \right)$, $\overline{\eta}_M = Q_{\text{Test}}^{-1} \sum_{\mu \in \Xi_{\text{Test}}^g} \eta_M(\mu)$, and $\varkappa_M$; here $\eta_M(\mu)$ is the effectivity defined in (11), and $\varkappa_M$ is the condition number of B^M. We present in Table 1 $\varepsilon_{M,\max}$, $\varepsilon_{M,\max}^*$, $\overline{\rho}_M$, Λ_M, $\overline{\eta}_M$, and $\varkappa_M$ as a function of M. We observe that $\varepsilon_{M,\max}$ and $\varepsilon_{M,\max}^*$ converge rapidly with M; that the Lebesgue constant provides a reasonably sharp measure of the interpolation-induced error; that the Lebesgue constant grows very slowly — and hence $\varepsilon_M(\mu)$ will be *only slightly larger than the min max result* $\varepsilon_M^*(\mu)$; that the error estimator effectivity is reasonably close to unity[3]; and that B^M is quite well-conditioned for our choice of basis. (For the non-orthogonalized basis ξ_m, $1 \le m \le M$, the condition number of B^M will grow exponentially with M.) These results are expected: although $G(x;\mu)$ varies rapidly as μ approaches 0 and x approaches 0, $G(x;\mu)$ is nevertheless quite smooth in the prescribed parameter domain $\mathcal{D}$.

[3]Note that the last column of the Table 1 in [4] (analogous to Table 1 here) contains an error — we purported in [4] to report the average of $\eta_M(\mu) \equiv \hat{\varepsilon}_M(\mu)/\varepsilon_M(\mu)$ over Ξ_{Test}^g, but in fact we reported the average over Ξ_{Test}^g of $\hat{\varepsilon}_M(\mu)/\varepsilon_M^*(\mu)$.

M	$\varepsilon_{M,\max}$	$\varepsilon^*_{M,\max}$	$\overline{\rho}_M$	Λ_M	$\overline{\eta}_M$	$\varkappa_M$
8	1.72 E−01	8.30 E−02	0.68	1.76	0.17	3.65
16	1.42 E−02	4.22 E−03	0.67	2.63	0.10	6.08
24	1.01 E−03	2.68 E−04	0.49	4.42	0.28	9.19
32	2.31 E−04	5.64 E−05	0.48	5.15	0.20	12.86
40	1.63 E−05	3.66 E−06	0.54	4.98	0.60	18.37
48	2.44 E−06	6.08 E−07	0.37	7.43	0.29	20.41

TABLE 1. Numerical results for empirical interpolation of $G(x;\mu)$: $\varepsilon_{M,\max}$, $\varepsilon^*_{M,\max}$, $\overline{\rho}_M$, Λ_M, $\overline{\eta}_M$, and $\varkappa_M$ as a function of M.

If we exploit the $L^2(\Omega)$-norm surrogate in our best approximation we can construct S^g_M much less expensively. We present in Table 2 numerical results obtained from this alternative construction of S^g_M. The results are very similar to those in Table 1, which implies — as expected — that the approximation quality of our empirical interpolation approach is relatively insensitive to the choice of norm in the sample construction process.

M	$\varepsilon_{M,\max}$	$\varepsilon^*_{M,\max}$	$\overline{\rho}_M$	Λ_M	$\overline{\eta}_M$	$\varkappa_M$
8	2.69 E−01	1.18 E−01	0.66	2.26	0.23	3.82
16	1.77 E−02	3.96 E−03	0.45	4.86	0.81	7.58
24	8.07 E−04	3.83 E−04	0.43	3.89	0.28	13.53
32	1.69 E−04	3.92 E−05	0.45	7.07	0.47	16.60
40	2.51 E−05	4.10 E−06	0.43	6.40	0.25	18.84
48	2.01 E−06	6.59 E−07	0.30	8.86	0.18	21.88

TABLE 2. Numerical results for empirical interpolation of $G(x;\mu)$: $\varepsilon_{M,\max}$, $\varepsilon^*_{M,\max}$, $\overline{\rho}_M$, Λ_M, $\overline{\eta}_M$, and $\varkappa_M$ as a function of M; here S^g_M is constructed with the $L^2(\Omega)$-norm as a surrogate for the $L^\infty(\Omega)$-norm.

3. Nonaffine Linear Coercive Elliptic Equations

3.1. **Problem Formulation**

3.1.1. *Abstract Statement*

We first define the Hilbert spaces $X^{\mathrm{e}} \equiv H^1_0(\Omega)$ — or, more generally, $H^1_0(\Omega) \subset X^{\mathrm{e}} \subset H^1(\Omega)$ — where $H^1(\Omega) = \{v \mid v \in L^2(\Omega), \nabla v \in (L^2(\Omega))^d\}$, $H^1_0(\Omega) = \{v \mid v \in H^1(\Omega), v|_{\partial\Omega} = 0\}$, and $L^2(\Omega)$ is the space of square integrable functions over Ω. The inner product and norm associated with X^{e} are given by $(\cdot,\cdot)_{X^{\mathrm{e}}}$ and $\|\cdot\|_{X^{\mathrm{e}}} = (\cdot,\cdot)^{1/2}_{X^{\mathrm{e}}}$, respectively; for example, $(w,v)_{X^{\mathrm{e}}} \equiv \int_\Omega \nabla w \cdot \nabla v + \int_\Omega w\, v,\ \forall w, v \in X^{\mathrm{e}}$. The truth approximation subspace X shall inherit this inner product and norm: $(\cdot;\cdot)_X \equiv (\cdot;\cdot)^{\mathrm{e}}_X$ and $\|\cdot\|_X \equiv \|\cdot\|^{\mathrm{e}}_X$.

In this section, we are interested in a particular form for problem (1), in which

$$a(w, v; \mu) = a_0(w, v) + a_1(w, v, g(\cdot; \mu)), \tag{13}$$

and

$$f(v; \mu) = \int_\Omega v h(x; \mu) \ , \tag{14}$$

where $a_0(\cdot, \cdot)$ is a (for simplicity, parameter-independent) bilinear form, $a_1 : X^{\rm e} \times X^{\rm e} \times L^\infty(\Omega)$ is a trilinear form, and $g(\cdot; \mu) \in L^\infty(\Omega)$, $h(\cdot; \mu) \in L^\infty(\Omega)$ are prescribed functions. For simplicity of exposition, we presume that $h(x; \mu) = g(x; \mu)$.

We shall assume that a satisfies coercivity and continuity conditions

$$0 < \alpha_0 \le \alpha(\mu) \equiv \inf_{w \in X \setminus \{0\}} \frac{a(w, w; \mu)}{\|w\|_X^2}, \quad \forall\, \mu \in \mathcal{D}, \tag{15}$$

$$\gamma(\mu) \equiv \sup_{w \in X \setminus \{0\}} \sup_{v \in X \setminus \{0\}} \frac{a(w, v; \mu)}{\|w\|_X \|v\|_X} \le \gamma_0 < \infty, \quad \forall\, \mu \in \mathcal{D}; \tag{16}$$

here $\alpha(\mu)$ and $\gamma(\mu)$ are the coercivity constant and the continuity constant, respectively. (We (plausibly) suppose that $\alpha_0,\ \gamma_0$ may be chosen independent of $\mathcal{N}$.) We shall further assume that the trilinear form a_1 satisfies

$$a_1(w, v, z) \le \gamma_1 \|w\|_X \, \|v\|_X \, \|z\|_{L^\infty(\Omega)}, \quad \forall\, w, v \in X, \forall z \in L^\infty(\Omega). \tag{17}$$

It is then standard, given that $g(\cdot; \mu) \in L^\infty(\Omega)$, to prove existence and uniqueness of the exact solution and the truth approximation.

3.1.2. *A Model Problem*

We consider the following model problem defined on the unit square $\Omega =]0, 1[^2 \in \mathbb{R}^2$: Given the parameter input $\mu = (\mu_{(1)}, \mu_{(2)}) \in \mathcal{D} \equiv [-1, -0.01]^2$, the field variable $u(\mu) \in X$ satisfies (2), where $X \subset X^{\rm e} \equiv H_0^1(\Omega)$ is a piecewise-linear finite element approximation space of dimension $\mathcal{N} = 2601$. Here a is given by (13) for

$$a_0(w, v) = \int_\Omega \nabla w \cdot \nabla v, \quad a_1(w, v, g(\cdot; \mu)) = \int_\Omega g(x; \mu) \ w \ v, \tag{18}$$

for $g(x; \mu) = G(x; \mu)$ as defined in (12); and f is given by (14) for $h(x; \mu) = g(x; \mu) = G(x; \mu)$. The output $s(\mu)$ is evaluated as $s(\mu) = \ell(u(\mu))$ for $\ell(v) = \int_\Omega \ v$.

The solution $u(\mu)$ develops a boundary layer in the vicinity of $x = (0,0)$ for μ near the "corner" $(-0.01, -0.01)$.

3.2. **Reduced-Basis Approximation**

3.2.1. *Discrete Equations*

We begin with motivating the need for the empirical interpolation approach in dealing with nonaffine problems; indeed, we shall continue the motivation discussed in Section 1. Specifically, we introduce the nested samples, $S_N^u = \{\mu_1^u \in \mathcal{D}, \cdots, \mu_N^u \in \mathcal{D}\},\ 1 \le N \le N_{\max}$, and associated nested Lagrangian[4] [28] reduced-basis spaces $W_N^u = \text{span}\{\zeta_j \equiv u(\mu_j^u),\ 1 \le j \le N\},\ 1 \le N \le N_{\max}$, where $u(\mu_j^u)$ is the solution of (2) for $\mu = \mu_j^u$. (In practice we orthonormalize the $\zeta_j, 1 \le j \le N$, with respect to $(\cdot,\cdot)_X$ so that $(\zeta_i, \zeta_j)_X = \delta_{ij}, 1 \le i, j \le N$; the resulting algebraic system will then be well-conditioned.)

Were we to follow the classical recipe, the reduced-basis approximation would be obtained by a standard Galerkin projection: given $\mu \in \mathcal{D}$, we evaluate $s_N(\mu) = \ell(u_N(\mu))$, where $u_N(\mu) \in W_N^u$ is the solution of

$$a_0(u_N(\mu), v) + a_1(u_N(\mu), v, g(\cdot;\mu)) = \int_\Omega g(x;\mu)v, \quad \forall v \in W_N^u. \tag{19}$$

If we now express $u_N(\mu) = \sum_{j=1}^N u_{N\,j}(\mu)\zeta_j$ and choose test functions $v = \zeta_n,\ 1 \le n \le N$, in (19), we obtain the $N \times N$ linear algebraic system

$$\sum_{j=1}^N \left(a_0(\zeta_i, \zeta_j) + a_1(\zeta_i, \zeta_j, g(\cdot;\mu))\right) u_{N\,j}(\mu) = \int_\Omega g(x;\mu)\zeta_i, \quad 1 \le i \le N. \tag{20}$$

We observe that while $a_0(\zeta_i, \zeta_j)$ is parameter-independent and can thus be pre-computed offline, $\int_\Omega g(x;\mu)\zeta_i$ and $a_1(\zeta_i, \zeta_j, g(\cdot;\mu))$ depend on $g(x;\mu)$ and must thus be evaluated online for every new parameter value μ; the operation count for the online stage will thus scale as $O(N^2\mathcal{N})$, where $\mathcal{N}$ is the dimension of the underlying truth finite element approximation space. The decrease in marginal cost in replacing the truth finite element approximation space with the reduced-basis approximation will be quite modest regardless of the dimension reduction $\mathcal{N} \to N \ll N$.

To recover online $\mathcal{N}$-independence, we appeal to the empirical interpolation method discussed in Section 2. We simply replace $g(x;\mu)$ in (20) with the (necessarily) affine approximation $g_M(x;\mu) = \sum_{m=1}^M \varphi_{M\,m}(\mu)q_m(x)$

[4] We may also consider Hermitian spaces built upon sensitivity derivatives of u with respect to μ [15] or, more generally, Lagrange-Hermitian spaces [17].

from (6) based upon the empirical interpolation approach described in Section 2. Our reduced-basis approximation is then: Given $\mu \in \mathcal{D}$, find $u_{N,M}(\mu) \in W_N^u$ such that

$$a_0(u_{N,M}(\mu), v) + a_1(u_{N,M}(\mu), v, g_M(\cdot;\mu)) = \int_\Omega g_M(x;\mu)v, \quad \forall v \in W_N^u; \tag{21}$$

we then evaluate the output estimate from

$$s_{N,M}(\mu) = \ell(u_{N,M}(\mu)). \tag{22}$$

We now express $u_{N,M}(\mu) = \sum_{j=1}^N u_{N,M\,j}(\mu)\,\zeta_j$, choose as test functions $v = \zeta_n,\ 1 \le n \le N$, and invoke (6) to obtain

$$\sum_{j=1}^N \left(a_0(\zeta_i,\zeta_j) + \sum_{m=1}^M \varphi_{M\,m}(\mu)\, a_1(\zeta_i,\zeta_j,q_m) \right) u_{N,M\,j}(\mu) = \sum_{m=1}^M \varphi_{M\,m}(\mu) \int_\Omega \zeta_i q_m, \quad 1 \le i \le N, \tag{23}$$

where $\varphi_{M\,m}(\mu),\ 1 \le m \le M$, is determined from (7). We indeed recover the online $\mathcal{N}$-independence: the quantities $a_0(\zeta_i,\zeta_j)$, $a_1(\zeta_i,\zeta_j,q_m)$, and $\int_\Omega \zeta_i q_m$ are all *parameter independent* and can thus be pre-computed offline, as discussed further in Section 3.2.3.

3.2.2. *A* Priori *Theory*

We consider here the convergence rate of $u_{N,M}(\mu) \to u(\mu)$. In fact, it is a simple matter to demonstrate the optimality of $u_{N,M}(\mu)$ in

Proposition 3.1. *For $\varepsilon_M(\mu)$ of (8) satisfying $\varepsilon_M(\mu) \le \frac{1}{2}\frac{\alpha(\mu)}{\phi_2(\mu)}$, we have*

$$\|u(\mu) - u_{N,M}(\mu)\|_X \le \left(1 + \frac{\gamma(\mu)}{\alpha(\mu)}\right) \inf_{w_N \in W_N^u} \|u(\mu) - w_N\|_X + \varepsilon_M(\mu) \left(\frac{\phi_1(\mu)\alpha(\mu) + 2\phi_2(\mu)\phi_3(\mu)}{\alpha^2(\mu)} \right) \ ; \tag{24}$$

here $\phi_1(\mu)$, $\phi_2(\mu)$, and $\phi_3(\mu)$ are given by

$$\phi_1(\mu) = \frac{1}{\varepsilon_M(\mu)} \sup_{v \in X} \frac{\int_\Omega v(g(x;\mu) - g_M(x;\mu))}{\|v\|_X} \ , \tag{25}$$

$$\phi_2(\mu) = \frac{1}{\varepsilon_M(\mu)} \sup_{w \in X} \sup_{v \in X} \frac{a_1(w,v;g(\cdot;\mu) - g_M(\cdot;\mu))}{\|w\|_X \|v\|_X} \ , \tag{26}$$

$$\phi_3(\mu) = \sup_{v \in X} \frac{\int_\Omega v g_M(x;\mu)}{\|v\|_X} \ . \tag{27}$$

Proof. For any $w_N = u_{N,M}(\mu) + v_N \in W_N^u$, we have

$$\begin{aligned}
\alpha(\mu)\|w_N - u_{N,M}(\mu)\|_X^2 &\leq a(w_N - u_{N,M}(\mu), w_N - u_{N,M}(\mu); \mu) \\
&= a(w_N - u(\mu), v_N; \mu) + a(u(\mu) - u_{N,M}(\mu), v_N; \mu) \\
&\leq \gamma(\mu)\|w_N - u(\mu)\|_X \|v_N\|_X + a(u(\mu) - u_{N,M}(\mu), v_N; \mu) \ . \qquad (28)
\end{aligned}$$

Note further from (2), (21), and (25)-(27) that the second term can be bounded by

$$\begin{aligned}
a(u(\mu) - u_{N,M}(\mu), v_N; \mu) &= \int_\Omega v_N g(x;\mu) - a(u_{N,M}(\mu), v_N; \mu) \\
&= \int_\Omega v_N (g(x;\mu) - g_M(x;\mu)) - a_1(u_{N,M}(\mu), v_N; g(x;\mu) - g_M(x;\mu)) \\
&\leq \varepsilon_M(\mu)\phi_1(\mu)\|v_N\|_X + \varepsilon_M(\mu)\phi_2(\mu)\|v_N\|_X \|u_{N,M}(\mu)\|_X \\
&\leq \varepsilon_M(\mu) \left(\frac{\phi_1(\mu)\alpha(\mu) + 2\phi_2(\mu)\phi_3(\mu)}{\alpha(\mu)} \right) \|v_N\|_X, \qquad (29)
\end{aligned}$$

where the last inequality derives from

$$\begin{aligned}
\alpha(\mu)\|u_{N,M}(\mu)\|_X^2 &\leq a(u_{N,M}(\mu), u_{N,M}(\mu); \mu) \\
&= \int_\Omega u_{N,M}(\mu) g_M(x;\mu) + a_1(u_{N,M}(\mu), u_{N,M}(\mu); g(x;\mu) - g_M(x;\mu)) \\
&\leq \phi_3(\mu)\|u_{N,M}(\mu)\|_X + \varepsilon_M(\mu)\phi_2(\mu)\|u_{N,M}(\mu)\|_X^2, \qquad (30)
\end{aligned}$$

and our hypothesis on $\varepsilon_M(\mu)$. It then follows from (28) and (29) that

$$\|w_N - u_{N,M}(\mu)\|_X \leq \frac{\gamma(\mu)}{\alpha(\mu)}\|w_N - u(\mu)\|_X + \varepsilon_M(\mu) \left(\frac{\phi_1(\mu)\alpha(\mu) + 2\phi_2(\mu)\phi_3(\mu)}{\alpha^2(\mu)} \right), \quad \forall\, w_N \in W_N^u \ . \qquad (31)$$

The desired result finally follows from (31) and the triangle inequality. (Note that ϕ_1, ϕ_2, and ϕ_3 are bounded by virtue of our continuity requirements.) □

We note from Proposition 3.1 that M should be chosen such that $\varepsilon_M(\mu)$ is of the same order as the error in the best approximation, $\inf_{w_N \in W_N^u} \|u(\mu) - w_N\|_X$, as otherwise the second term on the right-hand side of (24) may limit the convergence of the reduced-basis approximation. As regards the error in the best approximation, we note that W_N^u comprises "snapshots" on the parametrically induced manifold $\mathcal{M}^u \equiv \{u(\mu) \,|\, \forall\, \mu \in \mathcal{D}\} \subset X$. The critical observations are that $\mathcal{M}^u$ is very *low-dimensional* and that $\mathcal{M}^u$ is

smooth under general hypotheses on stability and continuity. We thus expect that the best approximation will converge to $u(\mu)$ very rapidly, and hence that N may be chosen small. (This is proven for a particularly simple case in [21].)

3.2.3. *Offline–Online Procedure*

We summarize here the procedure [3, 18, 20, 29]. In the offline stage — performed only *once* — we first construct nested approximation spaces W_M^g and nested sets of interpolation points T_M, $1 \le M \le M_{\max}$; we then choose S_N^u [5] and solve for (and orthonormalize) the ζ_n, $1 \le n \le N$; we finally form and store $a_0(\zeta_j, \zeta_i), a_1(\zeta_j, \zeta_i, q_m), \int_\Omega \zeta_i q_m$, and $\ell(\zeta_i), 1 \le i, j \le N, 1 \le m \le M_{\max}$. All quantities computed in the offline stage are independent of the parameter μ; note these quantities must be computed in a stable fashion which is consistent with the finite element quadrature points (see [23] page 173, and [12] page 132). In the online stage — performed many times for each new μ — we first compute $\varphi_M(\mu)$ from (7) at cost $O(M^2)$ by appealing to the triangular property of B^M; we then assemble and invert the (full) $N \times N$ reduced-basis stiffness matrix $a_0(\zeta_j, \zeta_i) + \sum \varphi_{M\,m}(\mu)\, a_1(\zeta_j, \zeta_i, q_m)$ to obtain $u_{N,M,\,j}, 1 \le j \le N$, at cost $O(N^2 M)$ for assembly plus $O(N^3)$ for inversion; we finally evaluate the reduced-basis output $s_{N,M}(\mu)$ as $s_{N,M}(\mu) = \sum_{j=1}^N u_{N,M,\,j}\ell(\zeta_j)$ at cost $O(N)$. The operation count for the online stage is thus only $O(M^2 + N^2 M + N^3)$.

Hence, as required in the many-query or real-time contexts, the online complexity is *independent* of $\mathcal{N}$, the dimension of the underlying "truth" finite element approximation space. Since $N,\ M \ll \mathcal{N}$ we expect significant computational savings in the online stage relative to classical discretization and solution approaches and relative to standard Galerkin reduced-basis approaches built upon (20).

3.2.4. *Numerical Results*

We present here numerical results for the model problem of Section 3.1.2. We first define $(w, v)_X = \int_\Omega \nabla w \cdot \nabla v$; thanks to the Dirichlet conditions on the boundary, $(w, v)_X$ is appropriately coercive. We note that for our particular function, $g(x; \mu) = G(x; \mu)$ of (12), S_M^g, W_M^g, and hence T_M and B^M are already

[5] In actual practice, our nested samples S_N^u and associated approximation spaces W_N^u are constructed by a greedy selection process [24, 29, 39] — which relies on *a posteriori* error estimators for the errors $\|u(\mu) - u_{N,M}(\mu)\|_X$ and $|s(\mu) - s_{N,M}(\mu)|$ — that ensures "maximally independent" snapshots and hence a rapidly convergent reduced-basis approximation. This sampling strategy, in conjunction with our orthogonalization procedure, also guarantees a well-conditioned reduced-basis discrete system [24, 29, 39]. Details of this sampling procedure and the *a posteriori* error estimation procedures can be found in [23] for elliptic problems and [12] for parabolic problems.

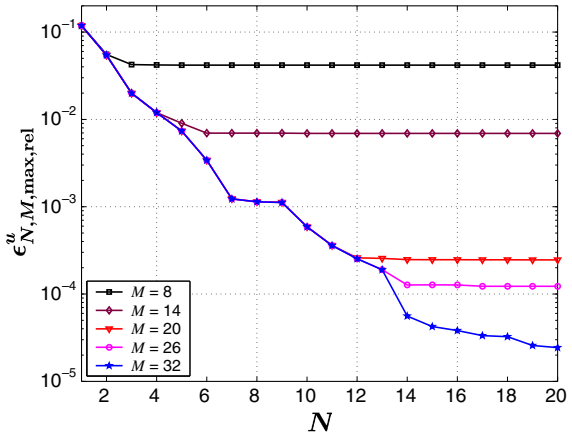


FIGURE 2. Convergence of the reduced-basis approximation for the nonaffine elliptic example.

constructed in Section 2.3. The sample set S^u_N and associated reduced-basis space W^u_N are developed based on the adaptive sampling procedure [23, 39] described in footnote 5.

We now introduce a parameter sample $\Xi_{\mathrm{Test}} \subset \mathcal{D}$ of size 225 (in fact, a regular 15×15 grid over $\mathcal{D}$), and define $\epsilon^u_{N,M,\max,\mathrm{rel}} = \max_{\mu\in\Xi_{\mathrm{Test}}} \|u(\mu) - u_{N,M}(\mu)\|_X / \|u_{\max}\|_X$ and $\epsilon^s_{N,M,\max,\mathrm{rel}} = \max_{\mu\in\Xi_{\mathrm{Test}}} |s(\mu) - s_{N,M}(\mu)|/|s_{\max}|$; here $\|u_{\max}\|_X = \max_{\mu\in\Xi_{\mathrm{Test}}} \|u(\mu)\|_X$ and $|s_{\max}| = \max_{\mu\in\Xi_{\mathrm{Test}}} |s(\mu)|$. We present in Figure 2 $\epsilon^u_{N,M,\max,\mathrm{rel}}$ as a function of N and M. We observe that the reduced-basis approximation converges very rapidly. We also note, consistent with Proposition 3.1, the "plateau" in the curves for M fixed and the "drops" in the $N \to \infty$ asymptotes as M is increased: for fixed M the error in our coefficient function approximation $g_M(x;\mu)$ to $g(x;\mu)$ will ultimately dominate for large N; increasing M renders the coefficient function approximation more accurate, which in turn leads to the drops in the asymptotic error. Figure 2 clearly suggests (for this particular problem) the optimal "$N - M$" strategy. We tabulate in Table 3 $\epsilon^u_{N,M,\max,\mathrm{rel}}$ and $\epsilon^s_{N,M,\max,\mathrm{rel}}$ for M chosen roughly optimally — but conservatively, to ensure that we are *not* on a "plateau" for each N. We observe very rapid convergence of the reduced-basis approximation with N, M. (Note that the convergence of the output can be further improved by the introduction of adjoint techniques [23, 24, 29].)

Finally, we present in Table 4 the online computational times to calculate $s_{N,M}(\mu)$ as a function of (N, M); the values are normalized with respect to the computational time for the direct calculation of the truth approximation output $s(\mu) = \ell(u(\mu))$. We achieve significant computational savings: for a relative

N	M	$\epsilon^u_{N,M,\text{max,rel}}$	$\epsilon^s_{N,M,\text{max,rel}}$
4	15	1.20 E – 02	5.96 E – 03
8	20	1.14 E – 03	2.42 E – 04
12	25	2.54 E – 04	1.76 E – 04
16	30	3.82 E – 05	7.92 E – 06

TABLE 3. Maximum relative error in the energy norm and output for the nonaffine elliptic example.

accuracy of close to 0.024 percent (corresponding to $N = 8$, $M = 20$ in Table 3) in the output, the online saving is more than a factor of 2000.

N	M	Online time for $s_{N,M}(\mu)$	(Online) time for $s(\mu)$
4	15	2.39 E – 04	1
8	20	4.33 E – 04	1
12	25	5.41 E – 03	1
16	30	6.93 E – 03	1

TABLE 4. Online computational times (normalized with respect to the time to solve for $s(\mu)$) for the nonaffine elliptic example.

4. Nonaffine Linear Parabolic Equations

4.1. **Problem Formulation**

We will now extend the results of the previous section to parabolic problems with nonaffine parameter dependence. The essential new ingredient is the presence of time; we shall "simply" treat time as an additional, albeit special, parameter. We note that we do not consider adjoint formulations for the parabolic problem in this paper — our primary focus here is on the treatment of the nonaffine and nonlinear terms. However, adjoint techniques can be gainfully employed for reduced-basis approximations of parabolic PDEs; see [14] for a detailed treatment of parabolic problems with affine parameter dependence by reduced-basis primal-dual approaches.

4.1.1. *Abstract Statement*

The "truth" finite element approximation is based on (3) for $Y^{\text{e}} \equiv L^2(\Omega)$; as in Section 3, a and f are of the form (13) and (14), respectively. We shall make the following assumptions. First, we assume that the bilinear form $a(\cdot,\cdot;\mu)$ is symmetric and satisfies the coercivity and continuity conditions (15) and (16), respectively. Second, we assume that the bilinear form $m(\cdot,\cdot)$ is symmetric $m(v,w) = m(w,v),\ \forall w, v \in Y^{\text{e}},\ \forall \mu \in \mathcal{D}$;

Y^{e}-coercive,

$$0 < \sigma \equiv \inf_{v \in Y^{\mathrm{e}}} \frac{m(v,v)}{\|v\|^2_{Y^{\mathrm{e}}}}, \quad \forall \mu \in \mathcal{D}; \tag{32}$$

and Y^{e}-continuous,

$$\sup_{w \in Y^{\mathrm{e}}} \sup_{v \in Y^{\mathrm{e}}} \frac{m(w,v)}{\|w\|_{Y^{\mathrm{e}}} \|v\|_{Y^{\mathrm{e}}}} \leq \rho < \infty, \quad \forall \mu \in \mathcal{D}. \tag{33}$$

(We (plausibly) suppose that ρ and σ may be chosen independent of $\mathcal{N}$.) We also require that the linear forms $f(\cdot;\mu) :\ X \to \mathbb{R}$ and $\ell(\cdot) :\ X \to \mathbb{R}$ be bounded with respect to $\|\cdot\|_{Y^{\mathrm{e}}}$; the former is perforce satisfied for the choice (14). Third, and finally, we assume that all linear and bilinear forms are independent of time — the system is thus linear time-invariant (LTI). It follows from our hypotheses that the finite element truth solution exists and is unique (see, e.g. [31]).

We note that the output and field variable are now functions of *both* the parameter μ and (discrete) time t^k. For simplicity of exposition, we assume here that m does not depend on the parameter; however, dependence on the parameter is readily admitted [14]. We also note that the method presented here easily extends to nonzero initial conditions, to multiple control inputs and outputs, and to nonsymmetric problems such as the convection-diffusion equation [12].

4.1.2. *Model Problem*

Our particular numerical example is the unsteady analog of the model problem introduced in Section 3.1.2: we recall that $\mu \in \mathcal{D} \equiv [-1, -0.01]^2$, that $\Omega =]0,1[^{\,2}$, and that our "truth" approximation subspace $X \equiv H^1_0(\Omega)$ is of dimension $\mathcal{N} = 2601$. The governing equation for $u(\mu, t^k) \in X$ is thus (3) with $a(w,v;\mu) = a_0(w,v) + a_1(w,v,G(\cdot;\mu))$, $f(v;\mu) = \int_\Omega v G(x;\mu)$,

$$m(w,v) \equiv \int_\Omega w\, v\ ; \tag{34}$$

recall that $G(x;\mu)$ is given by (12). The output is given by $s(\mu, t^k) = \ell(u(\mu, t^k)),\ \forall k \in \mathbb{K}$, where $\ell(v) = \int_\Omega v$. We shall consider the time interval $\bar{I} = [0,2]$ and a timestep $\Delta t = 0.01$; we thus have $K = 200$. Finally, we assume that we are given the periodic control input $b(t) = \sin(2\pi t),\ t \in \mathbb{I}$.

4.2. **Reduced-Basis Approximation**

4.2.1. *Fully Discrete Equations*

We first introduce the nested sample sets $S_N^u = \{\tilde{\mu}_1^u \in \tilde{\mathcal{D}}, \ldots, \tilde{\mu}_N^u \in \tilde{\mathcal{D}}\}$, $1 \le N \le N_{\max}$, where $\tilde{\mu} \equiv (\mu, t^k)$ and $\tilde{\mathcal{D}} \equiv \mathcal{D} \times \mathbb{I}$; note that the samples must now reside in the *parameter-time* space, $\tilde{\mathcal{D}}$. We then define the associated nested Lagrangian [28] reduced-basis space

$$W_N^u = \text{span}\{\zeta_n \equiv u(\tilde{\mu}_n^u),\ 1 \le n \le N\}, \quad 1 \le N \le N_{\max}, \tag{35}$$

where $u(\tilde{\mu}_n^u)$ is the solution of (3) at time $t = t^{k_n^u}$ for $\mu = \mu_n^u$. (As in the elliptic case, the ζ_n are orthonormalized relative to the $(\cdot;\cdot)_X$ inner product.)

Our reduced-basis approximation $u_{N,M}(\mu, t^k)$ to $u(\mu, t^k)$ is then obtained by a standard Galerkin projection: given $\mu \in \mathcal{D}$, $u_{N,M}(\mu, t^k) \in W_N^u$ satisfies

$$\begin{aligned} m(u_{N,M}(\mu,t^k),v) + \Delta t\ \left(a_0(u_{N,M}(\mu,t^k),v) + a_1(u_{N,M}(\mu,t^k),v;g_M(\cdot;\mu))\right) \\ = m(u_{N,M}(\mu,t^{k-1}),v) + \Delta t \int_\Omega v g_M(x;\mu)\ b(t^k), \quad \forall v \in W_N^u,\ \forall k \in \mathbb{K}, \end{aligned} \tag{36}$$

with initial condition $u_{N,M}(\mu, t^0) = 0$; here, $g_M(x;\mu)$ is the coefficient function approximation defined in (6). We then evaluate the output estimate, $s_{N,M}(\mu, t^k)$, from

$$s_{N,M}(\mu, t^k) \equiv \ell(u_{N,M}(\mu, t^k)), \quad \forall k \in \mathbb{K}. \tag{37}$$

The parameter-time sample set S_N^u and associated reduced-basis space W_N^u are constructed using a "greedy" adaptive sampling procedure summarized in footnote 5; we refer the interested reader to [14] for a detailed discussion of this procedure.

The reduced-basis subspace defined in (35) is the span of solutions of our "truth approximation" $u(\mu, t^k)$ at the sample points S_N^u. In many cases, however, the control input $b(t^k)$ is not known in advance and thus we cannot solve for $u(\mu, t^k)$ — as often arises in optimal control problems. Fortunately, we may appeal to the LTI hypotheses in such cases and construct the space based on the impulse response [14].

As regards the convergence rate $u_{N,M}(\mu, t^k) \longrightarrow u(\mu, t^k)$, we can develop a priori estimates very similar in form to the elliptic case — the sum of a best approximation result and a perturbation due to the variational

crime associated with the interpolation of g. The result is given in Proposition A.1 in the Appendix. It is also clear from Proposition A.1 that M should be chosen such that $\epsilon_M(\mu)$ is of the same order as the error in the best approximation, otherwise the perturbation term may limit the convergence of the reduced-basis approximation. As regards the best approximation, W_N^u comprises "snapshots" on the parametrically induced manifold $\mathcal{M}^u \equiv \{u(\mu, t^k) | \forall (\mu, t^k) \in \tilde{\mathcal{D}}\}$ which is very *low-dimensional* and *smooth* under general hypotheses on stability and continuity; the best approximation $u_{N,M}(\mu)$ should thus converge to $u(\mu, t^k)$ very rapidly.

The offline-online procedure for nonaffine linear parabolic equations is a straightforward combination of the procedures developed for affine parabolic equations [14] and nonaffine elliptic equations (see Section 3). For example, the online effort is $O(MN^2)$ to assemble the reduced-basis discrete system, $O(N^3 + KN^2)$ to obtain the reduced-basis coefficients at $t^k, 0 \le k \le K$, and $O(KN)$ to compute the output at $t^k, 0 \le k \le K$. (Recall that our system is LTI and hence the reduced-basis matrices are time-independent.)

4.2.2. *Numerical Results*

We now present numerical results for our model problem of Section 4.1.2. The sample set S_M^g and associated basis W_M^g — and hence T_M and B_M — for the nonaffine function approximation are constructed as in Section 2.3. We then generate the S_N^u and associated reduced-basis space W_N^u following the procedure of footnote 5; note for parabolic problems [12], we extract our snapshots from a parameter-*time* sample.

In the time-dependent case we define the maximum relative error in the energy norm as $\epsilon^u_{N,M,\max,\text{rel}} = \max_{\mu \in \Xi_{\text{Test}}} |||e(\mu, t^K)|||/|||u(\mu_u, t^K)|||$ and the maximum relative output error as $\epsilon^s_{N,M,\max,\text{rel}} = \max_{\mu \in \Xi_{\text{Test}}} |s(\mu, t_s(\mu)) - s_{N,M}(\mu, t_s(\mu))|/|s(\mu, t_s(\mu))|$. Here $\Xi_{\text{Test}} \subset \mathcal{D}$ is the parameter test sample of size 225 introduced in Section 3.2.4, $\mu_u \equiv \arg\max_{\mu \in \Xi_{\text{Test}}} |||u(\mu, t^K)|||$, $t_s(\mu) = \arg\max_{t^k \in \mathbb{I}} |s(\mu, t^k)|$, and the energy norm is defined as $|||v(\mu, t^k)||| = \left(m(v(\mu, t^k), v(\mu, t^k)) + \sum_{k'=1}^k a(v(\mu, t^{k'}), v(\mu, t^{k'}); g(\cdot; \mu))\ \Delta t \right)^{\frac{1}{2}}$, $\forall v \in L^\infty(\mathcal{D} \times \mathbb{I}; X)$. We plot in Figure 3 $\epsilon^u_{N,M,\max,\text{rel}}$ as a function of N and M. The graph shows the same behavior already observed in the elliptic case: the error levels off at smaller and smaller values as we increase M. In Table 5, we present $\epsilon^u_{N,M,\max,\text{rel}}$ and $\epsilon^s_{N,M,\max,\text{rel}}$ as a function of N and M; note that the tabulated (N, M) values correspond roughly to the optimal "knees" of the $N - M$-convergence curves. It is interesting to compare Table 3

N	M	$\epsilon^u_{N,M,\text{max,rel}}$	$\epsilon^s_{N,M,\text{max,rel}}$
5	8	4.12 E−02	4.23 E−02
10	16	3.12 E−03	3.03 E−03
20	24	1.97 E−04	1.79 E−04
30	32	2.46 E−05	7.65 E−06
40	40	4.27 E−06	2.21 E−06
50	48	7.48 E−07	1.29 E−07

TABLE 5. Maximum relative error in the energy norm and output for different values of N and M for the nonaffine parabolic problem.

(elliptic) and Table 5 (parabolic): as expected, for the same accuracy, the requisite M is roughly the same, since G is time-independent; however, N is larger for the parabolic case as u is a function of μ *and* time.

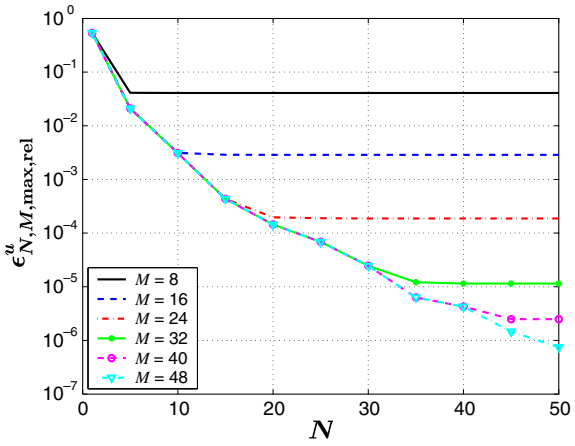


FIGURE 3. Convergence of the reduced-basis approximation for the non-affine parabolic example.

In Table 6 we present, as a function of N and M, the online computational times to calculate $s_{N,M}(\mu, t^k)$ and $\Delta^s_{N,M}(\mu, t^k)$, $\forall k \in \mathbb{K}$. The values are normalized with respect to the computational time for the direct calculation of the truth approximation output $s(\mu, t^k) = \ell(u(\mu, t^k))$, $\forall k \in \mathbb{K}$. The computational saving is quite significant: for a relative accuracy of roughly 0.02 percent ($N = 20$, $M = 24$) in the output, the online time to compute $s_{N,M}(\mu, t^k)$ is about 1/1000 the time to directly calculate $s(\mu, t^k)$.

5. NONLINEAR MONOTONIC ELLIPTIC EQUATIONS

5.1. **Problem Formulation**

5.1.1. *Abstract Statement*

Of course, nonlinear equations do not admit the same degree of generality as linear equations. We thus present our approach for a specific class of nonlinear equations. In particular, we consider the following

N	M	Online time for $s_{N,M}(\mu, t^k),\ \forall\, k \in \mathbb{K}$	(Online) time for $s(\mu, t^k),\ \forall\, k \in \mathbb{K}$
5	8	6.96 E−04	1
10	16	7.61 E−04	1
20	24	1.05 E−03	1
30	32	1.25 E−03	1
40	40	1.68 E−03	1
50	48	2.06 E−03	1

TABLE 6. Online computational times (normalized with respect to the time to solve for $s(\mu, t^k),\ \forall\, k \in \mathbb{K}$) for the nonaffine parabolic problem.

"exact" (superscript e) problem: for any $\mu \in \mathcal{D} \subset \mathbb{R}^P$, find $s^{\rm e}(\mu) = \ell(u^{\rm e}(\mu))$, where $u^{\rm e}(\mu) \in X^{\rm e}$ satisfies the weak form of the μ-parametrized nonlinear partial differential equation

$$a^L(u^{\rm e}(\mu), v) + \int_\Omega g(u^{\rm e}(\mu); x; \mu)v = f(v), \quad \forall\, v \in X^{\rm e}. \tag{38}$$

Here $g(u^{\rm e}; x; \mu)$ is a rather general nonaffine nonlinear function of the parameter μ, spatial coordinate x, and field variable $u^{\rm e}(x; \mu)$ (we present our assumptions later); and $a^L(\cdot, \cdot)$ and $f(\cdot), \ell(\cdot)$ are $X^{\rm e}$-continuous bounded bilinear and linear functionals, respectively — these forms are assumed to be parameter-independent for the sake of simplicity.

Next, we recall our reference (or "truth") finite element approximation space $X(\subset X^{\rm e})$ of dimension $\mathcal{N}$. Our truth approximation is then: given $\mu \in \mathcal{D}$, we find

$$s(\mu) = \ell(u(\mu)) \ , \tag{39}$$

where $u(\mu) \in X$ is the solution of the discretized weak formulation

$$a^L(u(\mu), v) + \int_\Omega g(u(\mu); x; \mu)v = f(v), \quad \forall v \in X \ . \tag{40}$$

We assume that $\|u^{\rm e}(\mu) - u(\mu)\|_X$ is suitably small and hence that $\mathcal{N}$ will typically be very large.

We shall make the following assumptions. First, we assume that the bilinear form $a^L(\cdot, \cdot) : X \times X \to \mathbb{R}$ is symmetric, $a^L(w, v) = a^L(v, w), \forall\, w, v \in X$. We shall also make two crucial hypotheses related to well-posedness. Our first hypothesis is that the bilinear form a^L satisfies a stability and continuity condition

$$0 < \alpha \equiv \inf_{v \in X} \frac{a^L(v, v)}{\|v\|_X^2} \ ; \tag{41}$$

$$\sup_{w\in X}\sup_{v\in X}\frac{a^L(w,v)}{\|w\|_X\|v\|_X}\equiv\gamma<\infty\ , \tag{42}$$

and that $f\in L^2(\Omega)$. For the second hypothesis we require that $g:\mathbb{R}\times\Omega\times\mathcal{D}\to\mathbb{R}$ is continuous in its arguments, increasing in its first argument, and satisfies, $\forall y\in\mathbb{R}$, $yg(y;x;\mu)\geq 0$ for any $x\in\Omega$ and $\mu\in\mathcal{D}$. With these assumptions, the problems (38) and (40) are indeed well-posed.

We can prove that there exists a solution $u^{\rm e}\in X^{\rm e}$ to the problem (38) first by considering the problem (38) with g replaced by

$$g_n(z;x;\mu)=\begin{cases} g(z;x;\mu) & \text{if} \quad |z|\leq n\\ -n & \text{if} \quad z<-n\\ n & \text{if} \quad z>n\end{cases} \tag{43}$$

and then taking the limit using Fatou's lemma (see [19]). In addition, the solution is unique: suppose indeed that (38) has two solution, $u_1^{\rm e}$ and $u_2^{\rm e}$; this implies

$$a^L(u_1^{\rm e}-u_2^{\rm e},v)+\int_\Omega\left(g(u_1;x;\mu)-g(u_2;x;\mu)\right)v=0,\quad\forall\, v\in H_0^1(\Omega)\ ;$$

by choosing $v=u_1^{\rm e}-u_2^{\rm e}$, we arrive at

$$a^L(u_1^{\rm e}-u_2^{\rm e},u_1^{\rm e}-u_2^{\rm e})+\int_\Omega\left(g(u_1;x;\mu)-g(u_2;x;\mu)\right)(u_1^{\rm e}-u_2^{\rm e})=0,\quad\forall\, v\in H_0^1(\Omega)\ ;$$

it follows from the coercivity of a^L and monotonicity of g in its first argument that $u_1^{\rm e}=u_2^{\rm e}$, and hence the solution is unique.

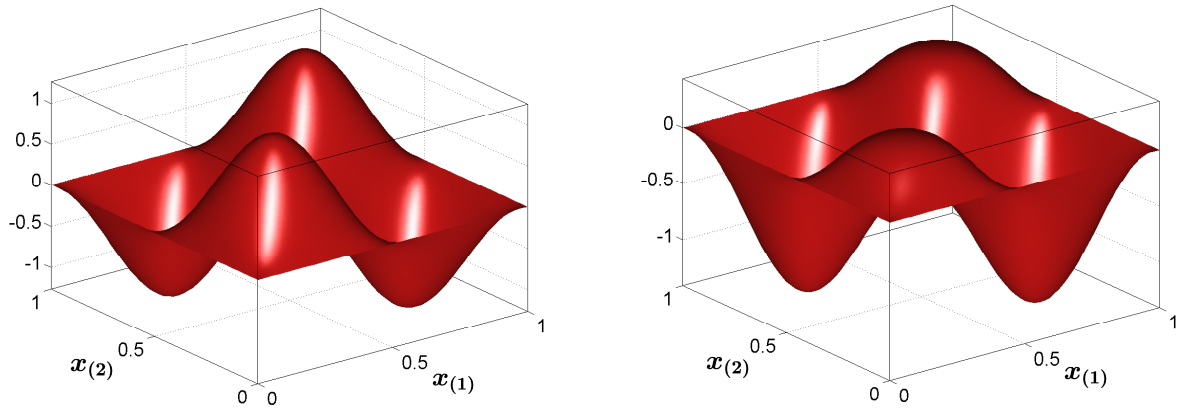
5.1.2. *A Model Problem*

We consider the model problem $-\nabla^2 u+\mu_{(1)}\frac{e^{\mu_{(2)}u}-1}{\mu_{(2)}}=100\sin(2\pi x_{(1)})\cos(2\pi x_{(2)})$, where $x_{(1)},x_{(2)}\in\Omega=]0,1[^2$ and $\mu=(\mu_{(1)},\mu_{(2)})\in\mathcal{D}^\mu\equiv[0.01,10]^2$; we impose a homogeneous Dirichlet condition on the boundary $\partial\Omega$. The output of interest is the average of the field variable over the physical domain. The weak formulation is then stated as: given $\mu\in\mathcal{D}$, find $s(\mu)=\int_\Omega u(\mu)$, where $u(\mu)\in X=H_0^1(\Omega)\equiv\{v\in H_1(\Omega)\mid v|_{\partial\Omega}=0\}$ is the solution of

$$\int_\Omega\nabla u\cdot\nabla v+\int_\Omega\mu_{(1)}\frac{e^{\mu_{(2)}u}-1}{\mu_{(2)}}\,v=100\int_\Omega\sin(2\pi x_{(1)})\,\cos(2\pi x_{(2)})\,v,\quad\forall v\in X\ . \tag{44}$$

Our abstract statement (39) and (40) then obtains for

$$a^L(w,v)=\int_\Omega\nabla w\cdot\nabla v,\quad f(v)=100\int_\Omega\sin(2\pi x_{(1)})\,\cos(2\pi x_{(2)})\,v,\quad \ell(v)=\int_\Omega v, \tag{45}$$



(a) (b)

FIGURE 4. Numerical solutions at typical parameter points for the nonlinear elliptic problem: (a) $\mu = (0.01, 0.01)$ and (b) $\mu = (10, 10)$.

and

$$g(y; x; \mu) = \mu_{(1)} \frac{e^{\mu_{(2)} y} - 1}{\mu_{(2)}}. \tag{46}$$

Note that $\mu_{(1)}$ controls the strength of the sink term and $\mu_{(2)}$ the strength of the nonlinearity. Clearly, g satisfies our hypotheses.

We present in Figure 4 two typical solutions obtained with the finite element "truth" approximation space X of dimension $\mathcal{N} = 2601$. We see that when $\mu = (0.01, 0.01)$, the solution has two negative peaks and two positive peaks with similar height (this solution is very similar to that of the problem in which $g(u; \mu)$ is zero). However, as μ increases, the negative peaks remain largely unchanged while the positive peaks are strongly rectified as shown in Figure 4(b) for $\mu = (10, 10)$: as μ increases the exponential function $\mu_{(1)} e^{\mu_{(2)} u}$ damps the positive part of $u(\mu)$, but has no effect on the negative part of $u(\mu)$.

5.2. **Reduced-Basis Approximation**

5.2.1. *Discrete Equations*

We first motivate the need for incorporating the empirical interpolation procedure into the reduced-basis method to treat nonlinear equations. If we were to directly apply the Galerkin procedure of the linear affine case, our reduced-basis approximation would satisfy

$$a^L(u_N(\mu), v) + \int_\Omega g(u_N(\mu); x; \mu) v = f(v), \quad \forall v \in W_N^u. \tag{47}$$

Observe that if g is a low order [18, 38] polynomial nonlinearity of u, we can then develop an efficient offline-online procedure. Unfortunately, this strategy can not be applied to high-order polynomial and non-polynomial nonlinearities: the operation count for the on-line stage will scale as some power of $\mathcal{N}$, the dimension of the truth finite element approximation space; the computational advantage relative to classical approaches using advanced iterative techniques is no longer obvious and in any event real-time response can not be guaranteed.

We seek an online evaluation cost that depends only on the dimension of reduced-basis approximation spaces and the parametric complexity of the problems — and *not* on $\mathcal{N}$. Towards that end, we first construct nested samples $S_M^g = \{\mu_1^g \in \mathcal{D}, \cdots, \mu_M^g \in \mathcal{D}\}$, associated nested approximation spaces $W_M^g = \text{span}\{\xi_m \equiv g(u(\mu_m^g); x; \mu_m^g), 1 \le m \le M\} = \text{span}\{q_1, \ldots, q_M\}$, and nested sets of interpolation points $T_M = \{x_1, \ldots, x_M\}$ for $1 \le M \le M_{\max}$ following the procedure of Section 2.1. Then for any given $w \in X$ and M, we approximate $g(w; x; \mu)$ by $g_M^w(x; \mu) = \sum_{m=1}^M \varphi_{M\,m}(\mu) q_m(x)$, where $\sum_{j=1}^M B_{i\,j}^M \varphi_{M\,j}(\mu) = g(w(x_i); x_i; \mu)$, $1 \le i \le M$.

We may now approximate $g(u_{N,M}; x; \mu)$ — as required in our reduced-basis projection for $u_{N,M}(\mu)$ — by $g_M^{u_{N,M}}(x; \mu)$. Our reduced-basis approximation is thus: Given $\mu \in \mathcal{D}$, we evaluate

$$s_{N,M}(\mu) = \ell(u_{N,M}(\mu)), \tag{48}$$

where $u_{N,M}(\mu) \in W_N^u$ satisfies

$$a^L(u_{N,M}(\mu), v; \mu) + \int_\Omega g_M^{u_{N,M}}(x; \mu) v = f(v), \quad \forall\, v \in W_N^u\ . \tag{49}$$

We now turn to the computational complexity.

5.2.2. *Offline–Online Procedure*

The most significant new issue is efficient calculation of the nonlinear term $g_M^{u_{N,M}}(x; \mu)$, which we now elaborate in some detail. We first expand our reduced-basis approximation and coefficient-function approximation as

$$u_{N,M}(\mu) = \sum_{j=1}^N u_{N,M\,j}(\mu) \zeta_j\ , \qquad g_M^{u_{N,M}}(x; \mu) = \sum_{m=1}^M \varphi_{M\,m}(\mu) q_m\ . \tag{50}$$

Inserting these representations into (49) yields

$$\sum_{j=1}^{N} A^N_{i\,j} u_{N,M\,j}(\mu) + \sum_{m=1}^{M} C^{N,M}_{i\,m} \varphi_{M\,m}(\mu) = F_{N\,i}, \quad 1 \le i \le N \ ; \tag{51}$$

where $A^N \in \mathbb{R}^{N\times N}, C^{N,M} \in \mathbb{R}^{N\times M}, F_N \in \mathbb{R}^N$ are given by $A^N_{i\,j} = a^L(\zeta_j, \zeta_i), 1 \le i,j \le N$, $C^{N,M}_{i\,m} = \int_\Omega q_m \zeta_i, 1 \le i \le N, 1 \le m \le M$, and $F_{N\,i} = f(\zeta_i), 1 \le i \le N$, respectively. Furthermore, $\varphi_M(\mu) \in \mathbb{R}^M$ is given by

$$\begin{aligned} \sum_{k=1}^{M} B^M_{m\,k} \varphi_{M\,k}(\mu) &= g(u_{N,M}(x_m;\mu); x_m; \mu), \quad 1 \le m \le M \\ &= g\Big(\sum_{n=1}^{N} u_{N,M\,n}(\mu)\zeta_n(x_m); x_m; \mu\Big), \quad 1 \le m \le M \ . \end{aligned} \tag{52}$$

We then substitute $\varphi_M(\mu)$ from (52) into (51) to obtain the following nonlinear algebraic system

$$\sum_{j=1}^{N} A^N_{i\,j} u_{N,M\,j}(\mu) + \sum_{m=1}^{M} D^{N,M}_{i\,m} g\Big(\sum_{n=1}^{N} \zeta_n(x_m) u_{N,M\,n}(\mu); x_m; \mu\Big) = F_{N\,i}, \quad 1 \le i \le N \ , \tag{53}$$

where $D^{N,M} = C^{N,M}(B^M)^{-1} \in \mathbb{R}^{N\times M}$.

To solve (53) for $u_{N,M\,j}(\mu), 1 \le j \le N$, we may apply a Newton iterative scheme: given a current iterate $\bar{u}_{N,M\,j}(\mu), 1 \le j \le N$, we find an increment $\delta u_{N,M\,j}, 1 \le j \le N$, such that

$$\begin{aligned} \sum_{j=1}^{N} \big(A^N_{i\,j} + \bar{E}^N_{i\,j}\big)\delta u_{N,M\,j}(\mu) = F_{N\,i} &- \sum_{j=1}^{N} A^N_{i\,j} \bar{u}_{N,M\,j}(\mu) \\ &- \sum_{m=1}^{M} D^{N,M}_{i\,m} g\Big(\sum_{n=1}^{N} \zeta_n(x_m)\bar{u}_{N,M\,n}(\mu); x_m; \mu\Big), \quad 1 \le i \le N \ ; \end{aligned} \tag{54}$$

here $\bar{E}^N \in \mathbb{R}^{N\times N}$ must be calculated at every Newton iteration as

$$\bar{E}^N_{i\,j} = \sum_{m=1}^{M} D^{N,M}_{i\,m} g_1\Big(\sum_{n=1}^{N} \zeta_n(x_m)\bar{u}_{N,M\,n}(\mu); x_m; \mu\Big)\zeta_j(x_m), \quad 1 \le i,j \le N \ , \tag{55}$$

where g_1 is the partial derivative of g with respect to its first argument. Finally, the output can be evaluated as

$$s_{N,M}(\mu) = \sum_{j=1}^{N} u_{N,M\,j}(\mu) L_{N\,j} \ , \tag{56}$$

where $L_N \in \mathbb{R}^N$ is the output vector with entries $L_{N\,j} = \ell(\zeta_j), 1 \le j \le N$. Based on this strategy, we can develop an efficient offline-online procedure for the rapid evaluation of $s_{N,M}(\mu)$ for each μ in $\mathcal{D}$.

The operation count of the online stage is essentially the predominant Newton update component (54): at each Newton iteration, we first assemble the right-hand side and compute $\bar{E}^N$ of (55) at cost $O(MN^2)$ — note we perform the sum in the parenthesis of (55) *before* performing the outer sum; we then form and invert the left-hand side (Jacobian) of (54) at cost $O(N^3)$. The online complexity depends only on N, M, and the number of Newton iterations; we thus recover $\mathcal{N}$ independence of the online stage.

5.2.3. *Numerical Results*

We first define $(w, v)_X = \int_\Omega \nabla w \cdot \nabla v$ and thus obtain $\alpha = 1$. We next construct S_M^g and W_M^g with the $L^2(\Omega)$-norm surrogate approach on Ξ^g, where Ξ^g is a regular 12×12 grid over $\mathcal{D}$. We then generate the sample set S_N^u and associated reduced-basis space W_N^u using the adaptive sampling construction [23] over the grid Ξ^g — but note for this nonlinear problem, our selection process is based directly on the energy norm of the true error (not an error estimate), $e(\mu) = u(\mu) - u_{N,M}(\mu)$, since the "truth" solutions $u(\mu)$ must be computed and stored for $\mu \in \Xi^g$ as part of the empirical interpolation procedure.

We now introduce a parameter test sample Ξ_{Test} of size 225 (a regular 15×15 grid) and define $\varepsilon^u_{N,M,\text{max,rel}} = \max_{\mu\in\Xi_{\text{Test}}} \|e_{N,M}(\mu)\|_X / \|u_{\text{max}}\|_X$ and $\varepsilon^s_{N,M,\text{max,rel}} = \max_{\mu\in\Xi_{\text{Test}}} |s(\mu) - s_{N,M}(\mu)|/|s_{\text{max}}|$, where $\|u_{\text{max}}\|_X = \max_{\mu\in\Xi_{\text{Test}}} \|u(\mu)\|_X$ and $|s_{\text{max}}| = \max_{\mu\in\Xi_{\text{Test}}} |s(\mu)|$; note that Ξ_{Test} is larger than (and mostly non-coincident with) Ξ^g. We present in Figure 5 $\varepsilon^u_{N,M,\text{max,rel}}$ as a function of N and M. We observe very rapid convergence of the reduced-basis approximation. Furthermore, the errors behave very similarly as in the linear example: the errors initially decrease, but then "plateau" in N for a particular value of M; increasing M effectively brings the error curves down. We also tabulate in Table 7 $\epsilon^u_{N,M,\text{max,rel}}$ and $\epsilon^s_{N,M,\text{max,rel}}$ for values of (N, M) close to the "knees" of the convergence curves of Figure 5. We see that $s_{N,M}(\mu)$ converges very rapidly.

N	M	$\epsilon^u_{N,M,\text{max,rel}}$	$\epsilon^s_{N,M,\text{max,rel}}$
4	5	6.53 E−03	2.11 E−02
8	10	1.05 E−03	2.38 E−03
12	15	7.34 E−05	1.26 E−04
16	20	1.30 E−05	2.79 E−05
20	25	5.05 E−06	8.00 E−06

TABLE 7. Maximum relative error in the energy norm and output for different values of (N, M) for the nonlinear elliptic problem.

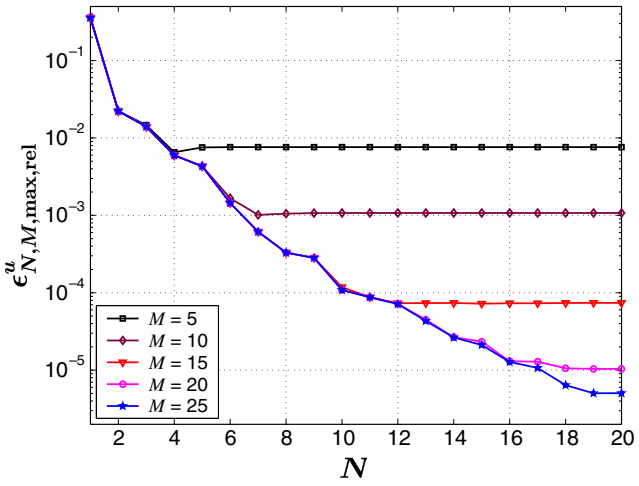


FIGURE 5. Convergence of the reduced-basis approximation for the nonlinear elliptic problem.

N	M	Online time for $s_{N,M}(\mu)$	(Online) time for $s(\mu)$
4	5	6.32 E−05	1
8	10	1.76 E−04	1
12	15	3.12 E−04	1
16	20	5.14 E−04	1
20	25	7.80 E−04	1

TABLE 8. Online computational times (normalized with respect to the time to solve for $s(\mu)$) for the nonlinear elliptic example.

We present in Table 8 the online computational times to calculate $s_{N,M}(\mu)$ as a function of (N, M). The values are normalized with respect to the computational time for the direct calculation of the truth approximation output $s(\mu) = \ell(u(\mu))$. The computational savings are much larger in the nonlinear case: for an relative accuracy of 0.0126 percent ($N = 12$, $M = 15$) in the output, the reduction in online cost is more than a factor of 3000; this is mainly because the matrix assembly of the nonlinear terms for the truth approximation is computationally very expensive. However we must also recall that, in the nonlinear case, the reduced-basis offline computations are much more extensive since we must solve the truth approximation over the large sample Ξ^g when constructing S^g_M.

6. Nonlinear Parabolic Equations

6.1. **Problem Formulation**

We now extend the results of the previous section to the time-dependent case and consider nonlinear parabolic problems. Similar to Section 4 we directly consider a time-discrete framework: we divide the time interval $\bar{I} \equiv [0, t_f]$ into K subintervals of equal length $\Delta t = \frac{t_f}{K}$ and define $t^k \equiv k\Delta t$, $0 \le k \le K \equiv \frac{t_f}{\Delta t}$, and $\mathbb{I} \equiv \{t^0, \ldots, t^k\}$; for the time integration we consider Euler-Backward. We also include a control input, $b(t^k)$, in the formulation of the problem.

6.1.1. *Abstract Statement*

We directly consider the "truth" approximation here. Our problem is based on the nonlinear elliptic problem (40) discussed in Section 5: Given a parameter $\mu \in \mathcal{D}$, we evaluate the output of interest

$$s(\mu, t^k) = \ell(u(\mu, t^k)), \quad \forall k \in \mathbb{K} \tag{57}$$

where the field variable $u(\mu, t^k) \in X, \ \forall k \in \mathbb{K}$,[6] satisfies the weak form of the nonlinear parabolic partial differential equation

$$m(u(\mu, t^k), v) + \Delta t \ a^L(u(\mu, t^k), v) + \Delta t \int_\Omega g(u(\mu, t^k); x; \mu)\, v = m(u(\mu, t^{k-1}), v) + \Delta t \ f(v) \ b(t^k), \quad \forall v \in X, \ \forall k \in \mathbb{K}, \tag{58}$$

with initial condition (say) $u(\mu, t^0) = 0$. (If an explicit scheme such as Euler-Forward is used, we then arrive at a linear system for $u(\mu, t^k)$ but now burdened with a conditional stability restriction on Δt. In that case, the discrete reduced-basis system is inheritedly linear.) We assume that a^L and m are symmetric and satisfy the coercivity conditions (41) and (32) and continuity conditions (42) and (33). We also require the linear forms $f(\cdot)$ and $\ell(\cdot)$ are bounded with respect to $\|\cdot\|_{Y^e}$. Since the focus of this section is the treatment of the nonlinearity $g(u(\mu, t); x; \mu)$, we assume for simplicity that m, a, f, and ℓ are parameter independent.

[6] We note that the field variable, $u(\mu, t^k)$, is of course also a function of the spatial coordinate x. In the sequel we will use the notation $u(x; \mu, t^k)$ to signify this dependence whenever it is crucial.

Similarly as in the steady case, assuming that $g : \mathbb{R}\times\Omega\times\mathcal{D} \to \mathbb{R}$ is continuous in its arguments, increasing in its first argument, and satisfies $\forall y \in \mathbb{R}, yg(y;x;\mu) \geq 0$ for any x and μ, it is a classical result of nonlinear analyses (truncation and monotonicity) to prove well-posedness of this problem (see [19]).

6.1.2. *Model Problem*

Our particular numerical example is the unsteady analog of the elliptic model problem introduced in Section 5.1.2: we have $\mu = (\mu_{(1)}, \mu_{(2)}) \in \mathcal{D}^\mu \equiv [0.01, 10]^2$, the spatial domain is the unit square, $\Omega =]0,1[^2$, and our "truth" approximation finite element space $X = H_0^1(\Omega)$ has dimension $\mathcal{N} = 2601$. The field variable $u(\mu, t^k) \in X$ thus satisfies (58) with

$$m(v,w) \equiv \int_\Omega v\, w, \qquad a^L(v,w) \equiv \int_\Omega \nabla v \cdot \nabla w, \qquad f(v) \equiv 100 \int_\Omega v\, \sin(2\pi x_{(1)})\, \cos(2\pi x_{(2)}), \tag{59}$$

and

$$g(u(\mu,t^k);\mu) = \mu_{(1)}\, \frac{e^{\mu_{(2)}\, u(\mu,t^k)} - 1}{\mu_{(2)}}. \tag{60}$$

The output $s(\mu, t^k)$ is evaluated from (57) with $\ell(v) = \int_\Omega v$. We shall consider the time interval $\bar{I} = [0, 2]$ and a timestep $\Delta t = 0.01$; we thus have $K = 200$. The control input is given by $b(t^k) = \sin(2\pi t^k),\ t \in \mathbb{I}$.

6.2. **Reduced-Basis Approximation**

6.2.1. *Fully Discrete Equations*

We first introduce the nested sample sets $S_M^g = \{\tilde{\mu}_1^g \in \tilde{\mathcal{D}}, \ldots, \tilde{\mu}_M^g \in \tilde{\mathcal{D}}\}$, $1 \leq M \leq M_{\max}$ and $S_N^u = \{\tilde{\mu}_1^u \in \tilde{\mathcal{D}}, \ldots, \tilde{\mu}_N^u \in \tilde{\mathcal{D}}\}$, $1 \leq N \leq N_{\max}$, where $\tilde{\mu} \equiv (\mu, t^k)$ and $\tilde{\mathcal{D}} \equiv \mathcal{D} \times \mathbb{I}$. Note that, since $g(\cdot;x;\mu)$ is a function of the field variable $u(\mu, t^k)$, the sample set S_M^g must now also reside in *parameter-time* space $\tilde{\mathcal{D}}$; in general, $S_N^u \neq S_M^g$ and in fact $N \neq M$. We define the nested collateral reduced-basis space

$$W_M^g = \text{span}\{\xi_n \equiv g(u(\tilde{\mu}_n^g);x;\mu),\ 1 \leq n \leq M\} = \text{span}\{q_1, \ldots,\ q_M\}, \quad 1 \leq M \leq M_{\max}, \tag{61}$$

and nested set of interpolation points $T_M = \{x_1, \ldots, x_M\},\ 1 \leq M \leq M_{\max}$; here $u(\tilde{\mu}_n^g)$ is the solution of (58) at time $t = t^{k_n^g}$ for $\mu = \mu_n^g$. Next, we define the associated nested Lagrangian [28] reduced-basis space

$$W_N^u = \text{span}\{\zeta_n \equiv u(\tilde{\mu}_n^u),\ 1 \leq n \leq N\}, \quad 1 \leq N \leq N_{\max}, \tag{62}$$

where $u(\tilde{\mu}_n^u)$ is the solution of (58) at time $t = t^{k_n^u}$ for $\mu = \mu_n^u$.

Our reduced-basis approximation $u_{N,M}(\mu, t^k)$ to $u(\mu, t^k)$ is then given by: given $\mu \in \mathcal{D}$, $u_{N,M}(\mu, t^k) \in W_N^u$ satisfies

$$m(u_{N,M}(\mu,t^k),v) + \Delta t\ a^L(u_{N,M}(\mu,t^k),v) + \Delta t\ \int_\Omega g_M^{u_{N,M}}(x;\mu,t^k)\ v \\ = m(u_{N,M}(\mu,t^{k-1}),v) + \Delta t\ f(v)\ b(t^k), \quad \forall v \in W_N^u,\ \forall k \in \mathbb{K}, \quad (63)$$

with initial condition $u_{N,M}(\mu, t^0) = 0$; here, $g_M^{u_{N,M}}(x;\mu,t^k)$ is the approximation to $g(u_{N,M}(\mu,t^k);x;\mu)$ given by

$$g_M^{u_{N,M}}(x;\mu,t^k) = \sum_{m=1}^{M} \varphi_{Mm}(\mu,t^k)\ q_m(x) \quad (64)$$

where the coefficients $\varphi_{Mm}(\mu, t^k)$ are determined from

$$\sum_{j=1}^{M} B_{ij}^M\ \varphi_{Mj}(\mu,t^k) = g(u_{N,M}(x_i;\mu,t^k);x_i;\mu), \quad 1 \le i \le M, \quad (65)$$

and $B_{ij}^M = q_j(x_i),\ 1 \le i,j \le M$. Finally, we evaluate the output from

$$s_{N,M}(\mu,t^k) = \ell(u_{N,M}(\mu,t^k)), \quad \forall k \in \mathbb{K}. \quad (66)$$

(Note that, contrary to the previous sections, $\varphi_M(\mu, t^k)$ now also depends on time.)

At this point we should remark that our current approach of constructing the sample set S_M^g and associated reduced-basis space W_M^g in the nonlinear parabolic case is computationally very expensive. The reason, related to our greedy adaptive sampling procedure proposed in Section 2.1, is twofold. First, we need to calculate and store the "truth" solution $u(\mu, t^k)$ at all times $t^k \in \mathbb{I}$ on the grid Ξ^g in parameter space. In our numerical example Ξ^g is of size 144 — we thus need to solve (58) 144 times and store 144×200 "truth" solutions $u(\mu, t^k)$! Second, as pointed out in Section 2.1, to determine the next sample point $\tilde{\mu}_n^g$ in $\tilde{\Xi}^g \equiv \Xi^g \times \mathbb{I}$, requires the solution of a linear program for all $\mu \in \tilde{\Xi}^g$ if g is time-varying — as is inherently the case in the nonlinear context.[7] Since this computation is too expensive in our current implementation, we revert to the least squares surrogate in this section — in choosing this approach we in fact rely on our numerical comparison in Section 2.1 showing that we can expect similar results.

[7]Note that in the linear nonaffine parabolic case the function g depended only on x and μ and *not* on time.

6.2.2. *Offline–Online Procedure*

The offline-online decomposition follows directly from the corresponding procedures for linear nonaffine parabolic problems (see Section 4) and nonlinear elliptic problems (see Section 5). In summary, the operation count (per Newton iteration per timestep) in the online stage is $O(MN^2 + N^3)$; the system is of course no longer "LTI".

We remark that, in actual practice, M can be quite large — and in fact much larger than N. We can reduce M without sacrificing accuracy by splitting the time interval $\mathbb{I}$ into several smaller subintervals $\mathbb{I}_1, \ldots, \mathbb{I}_{\mathcal{I}}$ such that $\mathbb{I} = \bigcup_{i=1,\mathcal{I}} \mathbb{I}_i$. We then construct, in the offline stage, $\mathcal{I}$ separate samples sets $S^g_{M\,i},\ 1 \le i \le \mathcal{I}$, and associated reduced-basis spaces $W^g_{M\,i},\ 1 \le i \le \mathcal{I}$, on each interval $\mathbb{I}_i,\ 1 \le i \le \mathcal{I}$. In the online stage we simply "switch" to the corresponding sample — and hence T_M, B^M, and $D^{N,M}$ — as time progresses. This approach renders the offline computation more expensive (and online storage more extensive), but can increase the online efficiency considerably while retaining the desired accuracy.

6.2.3. *Numerical Results*

We now present numerical results for our model problem of Section 6.1.2. We construct S^g_M and hence W^g_M with the surrogate least squares approach on $\tilde{\Xi}^g = \Xi^g \times \mathbb{I}$, where Ξ^g is a regular 12×12 grid over $\mathcal{D}$. We generate the sample set S^u_N and associated reduced-basis space W^u_N using an adaptive sampling procedure — but note for this nonlinear parabolic problem, our selection process is based directly on the energy norm of the true error (not an error estimate), $e(\mu, t^k) = u(\mu) - u_{N,M}(\mu, t^k)$, since the "truth" solutions $u(\mu, t^k)$ are stored for $\mu \in \Xi^g$.

We now define the maximum relative error in the energy norm $\epsilon^u_{N,M,\max,\mathrm{rel}} = \max_{\mu\in\Xi_{\mathrm{Test}}} |||e(\mu, t^K)|||/|||u(\mu_u, t^K)|||$ and the maximum relative output error $\epsilon^s_{N,M,\max,\mathrm{rel}} = \max_{\mu\in\Xi_{\mathrm{Test}}} |s(\mu, t_s(\mu)) - s_{N,M}(\mu, t_s(\mu))|/|s(\mu, t_s(\mu))|$. Here $\Xi_{\mathrm{Test}} \subset \mathcal{D}$ is the parameter test sample of size 225 introduced in Section 5.2.3, $\mu_u \equiv \arg\max_{\mu\in\Xi_{\mathrm{Test}}} |||u(\mu, t^K)|||$, $t_s(\mu) = \arg\max_{t^k\in\mathbb{I}} |s(\mu, t^k)|$, and the energy norm is defined as $|||v(\mu, t^k)||| \equiv \big(m(v(\mu, t^k), v(\mu, t^k)) + \sum_{k'=1}^{k} a^L(v(\mu, t^{k'}), v(\mu, t^{k'}))\ \Delta t\ \big)^{\frac{1}{2}}$, $\forall v \in L^\infty(\mathcal{D} \times \mathbb{I}; X)$.

We plot in Figure 6 $\epsilon^u_{N,M,\max,\mathrm{rel}}$ as a function of N for different values of M. We observe the same behavior as in the nonlinear elliptic case. We note, however, that M is now much larger compared to the

N	M	$\epsilon^y_{N,M,\max,\mathrm{rel}}$	$\epsilon^s_{N,M,\max,\mathrm{rel}}$
1	10	3.82 E – 01	1.00 E – 00
5	30	1.36 E – 02	1.91 E – 02
10	50	1.62 E – 03	1.46 E – 04
20	80	1.46 E – 04	1.67 E – 05
30	110	1.88 E – 05	5.16 E – 06
40	140	4.94 E – 06	1.56 E – 06

TABLE 9. Relative error in the energy norm and output for the nonlinear parabolic problem.

nonlinear elliptic model problem due to the time dependence; we recall that in the linear nonaffine elliptic and parabolic cases the required M was the same since the nonaffine coefficient function did not depend on time. In Table 9, we present $\epsilon^u_{N,M,\max,\mathrm{rel}}$ and $\epsilon^s_{N,M,\max,\mathrm{rel}}$ as a function of N and M. We observe very rapid convergence of the reduced-basis approximation; for $N = 20$ and $M = 80$ the error in the output is less than one percent.

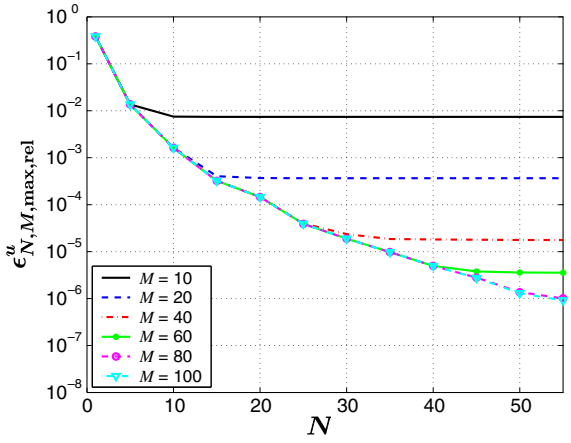


FIGURE 6. Convergence of the reduced-basis approximation for the nonlinear parabolic problem.

In Table 10 we present, as a function of N and M, the online computational times to calculate $s_{N,M}(\mu, t^k)$ and $\Delta^s_{N,M}(\mu, t^k)$, $\forall k \in \mathbb{K}$. The values are normalized with respect to the computational time for the direct calculation of the truth approximation output $s(\mu, t^k) = \ell(u(\mu, t^k))$, $\forall k \in \mathbb{K}$. The reduction in online response time is considerable. We again caution that the offline computations necessary in the nonlinear case are very extensive — primarily due to the sampling procedure for S^q_M. However, if a many-query context, or a clear demand for real-time response, can justify the offline cost, the reduced-basis methods can be very gainfully employed.

N	M	Online time for $s_{N,M}(\mu, t^k),\ \forall\, k \in \mathbb{K}$	(Online) time for $s(\mu, t^k),\ \forall\, k \in \mathbb{K}$
1	10	6.62 E – 05	1
5	30	1.19 E – 04	1
10	50	1.74 E – 04	1
20	80	3.88 E – 04	1
30	110	7.20 E – 04	1
40	140	1.22 E – 03	1

TABLE 10. Online computational times (normalized with respect to the time to solve for $s(\mu, t^k),\ \forall k \in \mathbb{K}$) for the nonlinear parabolic problem.

This work was supported by DARPA and AFOSR under Grant FA9550-05-1-0114 and by the Singapore-MIT Alliance. We would like to thank Maxime Barrault, Éric Cancès, Claude Le Bris, Gabriel Turinici, George Pau, and Sugata Sen for many stimulating and beneficial exchanges.

REFERENCES

[1] B. O. Almroth, P. Stern, and F. A. Brogan. Automatic choice of global shape functions in structural analysis. *AIAA Journal*, 16:525–528, May 1978.

[2] Z. J. Bai. Krylov subspace techniques for reduced-order modeling of large-scale dynamical systems. *Applied Numerical Mathematics*, 43(1-2):9–44, October 2002.

[3] E. Balmes. Parametric families of reduced finite element models: Theory and applications. *Mechanical Systems and Signal Processing*, 10(4):381–394, 1996.

[4] M. Barrault, N. C. Nguyen, Y. Maday, and A. T. Patera. An "empirical interpolation" method: Application to efficient reduced-basis discretization of partial differential equations. *C. R. Acad. Sci. Paris, Série I.*, 339:667–672, 2004.

[5] A. Barrett and G. Reddien. On the reduced basis method. *Z. Angew. Math. Mech.*, 75(7):543–549, 1995.

[6] T.T. Bui, M. Damodaran, and K. Wilcox. Proper orthogonal decomposition extensions for parametric applications in transonic aerodynamics (AIAA Paper 2003-4213). In *Proceedings of the 15th AIAA Computational Fluid Dynamics Conference*, June 2003.

[7] J. Chen and S-M. Kang. Model-order reduction of nonlinear mems devices through arclength-based karhunen-loéve decomposition. In *Proceeding of the IEEE international Symposium on Circuits and Systems*, volume 2, pages 457–460, 2001.

[8] Y. Chen and J. White. A quadratic method for nonlinear model order reduction. In *Proceeding of the international Conference on Modeling and Simulation of Microsystems*, pages 477–480, 2000.

[9] E.A. Christensen, M. Brøns, and J.N. Sørensen. Evaluation of proper orthogonal decomposition-based decomposition techniques applied to parameter-dependent nonturbulent flows. *SIAM J. Scientific Computing*, 21(4):1419–1434, 2000.

[10] P. Erdös. Problems and results on the theory of interpolation, ii. *Acta Math. Acad. Sci.*, 12:235–244, 1961.

[11] J. P. Fink and W. C. Rheinboldt. On the error behavior of the reduced basis technique for nonlinear finite element approximations. *Z. Angew. Math. Mech.*, 63:21–28, 1983.

[12] M. Grepl. *Reduced-Basis Approximations for Time-Dependent Partial Differential Equations: Application to Optimal Control*. PhD thesis, Massachusetts Institute of Technology, 2005.

[13] M. A. Grepl, N. C. Nguyen, K.Veroy, A. T. Patera, and G. R. Liu. Certified rapid solution of parametrized partial differential equations for real-time applications. In *Proceedings of the 2$^{\text{nd}}$ Sandia Workshop of PDE-Constrained Optimization: Towards Real-Time and On-Line PDE-Constrained Optimization*, SIAM Computational Science and Engineering Book Series, 2004. Submitted.

[14] M. A. Grepl and A. T. Patera. *A Posteriori* error bounds for reduced-basis approximations of parametrized parabolic partial differential equations. *M2AN (Math. Model. Numer. Anal.)*, 2004. To appear.

[15] Ph. Guillaume and M. Masmoudi. Solution to the time-harmonic maxwell's equations in a waveguide: use of higher-order derivatives for solving the discrete problem. *SIAM J. Numer. Anal.*, 34(4):1306–1330, 1997.

[16] M. D. Gunzburger. *Finite Element Methods for Viscous Incompressible Flows: A Guide to Theory, Practice, and Algorithms*. Academic Press, Boston, 1989.

[17] K. Ito and S. S. Ravindran. A reduced basis method for control problems governed by PDEs. In W. Desch, F. Kappel, and K. Kunisch, editors, *Control and Estimation of Distributed Parameter Systems*, pages 153–168. Birkhäuser, 1998.

[18] K. Ito and S. S. Ravindran. A reduced-order method for simulation and control of fluid flows. *Journal of Computational Physics*, 143(2):403–425, July 1998.

[19] J.L. Lions. *Quelques Methods de Resolution des Problemes aux Limites Non-lineares*. Dunod, 1969.

[20] L. Machiels, Y. Maday, I. B. Oliveira, A. T. Patera, and D. V. Rovas. Output bounds for reduced-basis approximations of symmetric positive definite eigenvalue problems. *C. R. Acad. Sci. Paris, Série I*, 331(2):153–158, July 2000.
[21] Y. Maday, A. T. Patera, and G. Turinici. Global *a priori* convergence theory for reduced-basis approximation of single-parameter symmetric coercive elliptic partial differential equations. *C. R. Acad. Sci. Paris, Série I*, 335(3):289–294, 2002.
[22] M. Meyer and H. G. Matthies. Efficient model reduction in non-linear dynamics using the karhunen-loève expansion and dual-weighted-residual methods. *Computational Mechanics*, 31(1-2):179–191, May 2003.
[23] N. C. Nguyen. *Reduced-Basis Approximation and* A Posteriori *Error Bounds for Nonaffine and Nonlinear Partial Differential Equations: Application to Inverse Analysis*. PhD thesis, Singapore-MIT Alliance, National University of Singapore., 2005.
[24] N. C. Nguyen, K. Veroy, and A. T. Patera. Certified real-time solution of parametrized partial differential equations. In *Handbook of Materials Modeling*. Kluwer Academic Publishing, 2004. To appear.
[25] A. K. Noor and J. M. Peters. Reduced basis technique for nonlinear analysis of structures. *AIAA Journal*, 18(4):455–462, April 1980.
[26] J. S. Peterson. The reduced basis method for incompressible viscous flow calculations. *SIAM J. Sci. Stat. Comput.*, 10(4):777–786, July 1989.
[27] J.R. Phillips. Projection-based approaches for model reduction of weakly nonlinear systems, time-varying systems. In *IEEE Transactions On Computer-Aided Design of Integrated Circuit and Systems*, volume 22, pages 171–187, 2003.
[28] T. A. Porsching. Estimation of the error in the reduced basis method solution of nonlinear equations. *Mathematics of Computation*, 45(172):487–496, October 1985.
[29] C. Prud'homme, D. Rovas, K. Veroy, Y. Maday, A. T. Patera, and G. Turinici. Reliable real-time solution of parametrized partial differential equations: Reduced-basis output bound methods. *Journal of Fluids Engineering*, 124(1):70–80, March 2002.
[30] A. Quarteroni, R. Sacco, and F. Saleri. *Numerical Mathematics*, volume 37 of *Texts in Applied Mathematics*. Springer, New York, 1991.
[31] A. Quarteroni and A. Valli. *Numerical Approximation of Partial Differential Equations*. Springer, 2nd edition, 1997.
[32] M. Rewienski and J. White. A trajectory piecewise-linear approach to model order reduction and fast simulation of nonlinear circuits and micromachined devices. In *IEEE Transactions On Computer-Aided Design of Integrated Circuit and Systems*, volume 22, pages 155–170, 2003.
[33] W. C. Rheinboldt. On the theory and error estimation of the reduced basis method for multi-parameter problems. *Nonlinear Analysis, Theory, Methods and Applications*, 21(11):849–858, 1993.
[34] T. J. Rivlin. *An introduction to the approximation of functions*. Dover Publications Inc., New York, 1981.
[35] J. M. A. Scherpen. Balancing for nonlinear systems. *Systems and Control Letters*, 21:143–153, 1993.
[36] L. Sirovich. Turbulence and the dynamics of coherent structures, part 1: Coherent structures. *Quarterly of Applied Mathematics*, 45(3):561–571, October 1987.
[37] S. Sugata. *Reduced Basis Approximation and A Posteriori Error Estimation for Many-Parameter Problems*. PhD thesis, Massachusetts Institute of Technology, 2007. In progress.
[38] K. Veroy and A. T. Patera. Certified real-time solution of the parametrized steady incompressible navier-stokes equations; Rigorous reduced-basis *a posteriori* error bounds. *International Journal for Numerical Methods in Fluids*, 2004. To appear.
[39] K. Veroy, C. Prud'homme, D. V. Rovas, and A. T. Patera. *A posteriori* error bounds for reduced-basis approximation of parametrized noncoercive and nonlinear elliptic partial differential equations (AIAA Paper 2003-3847). In *Proceedings of the 16th AIAA Computational Fluid Dynamics Conference*, June 2003.
[40] K. Veroy, D. Rovas, and A. T. Patera. *A Posteriori* error estimation for reduced-basis approximation of parametrized elliptic coercive partial differential equations: "Convex inverse" bound conditioners. *Control, Optimisation and Calculus of Variations*, 8:1007–1028, June 2002. Special Volume: A tribute to J.-L. Lions.
[41] D.S. Weile, E. Michielssen, and K. Gallivan. Reduced-order modeling of multiscreen frequency-selective surfaces using Krylov-based rational interpolation. *IEEE Transactions on Antennas and Propagation*, 49(5):801–813, May 2001.

Appendix

We consider here the rate at which $u_{N,M}(\mu, t^k)$ converges to $u(\mu, t^k)$ for the nonaffine linear parabolic case. As for the elliptic case, the interpolation-induced error will be measured through the functions $\phi_1(\mu), \phi_2(\mu)$ and $\phi_3(\mu)$ of (25)-(27) together with a comparison with respect to some best fit of $u(\mu, .)$ by elements of W_N^u. The natural measure for the best fit is the "$m + \Delta ta$" norm. We thus introduce the projector π_N defined by

$$m(v - \pi_N(v), w_N) + \Delta ta(v - \pi_N(v), w_N) = 0, \quad \pi_N(v) \in W_N^u, \forall w_N \in W_N^u, \forall v \in X \ .$$

We can then prove

Proposition 6.1. *For $\varepsilon_M(\mu)$ of (8) satisfying $\varepsilon_M(\mu) < \alpha(\mu)/(4\,\phi_2(\mu))$ (say), the error $e(\mu, t^k) \equiv u(\mu, t^k) - u_{N,M}(\mu, t^k)$ satisfies*

$$\sigma \left\| e(\mu, t^k) \right\|_Y^2 + \frac{\alpha(\mu)}{2}\, \Delta t \sum_{k'=1}^{k} \left\| e(\mu, t^{k'}) \right\|_X^2 \le \Upsilon(\mu)\, \Delta t \sum_{k'=1}^{k} b(t^{k'})^2 + 8\rho \left\| u(\mu, t^k) - \pi_N[u(\mu, t^k)] \right\|_Y^2 + (8\gamma(\mu) + 4\,\alpha(\mu)) \Delta t \sum_{k'=1}^{k} \left\| u(\mu, t^{k'}) - \pi_N[u(\mu, t^{k'})] \right\|_X^2, \quad \text{(A.1)}$$

where

$$\Upsilon(\mu) = \frac{18}{\alpha(\mu)}\, \varepsilon_M(\mu)^2 \left(\phi_1(\mu)^2 + \phi_2(\mu)^2 \frac{2\,\phi_3(\mu)^2}{\alpha(\mu)^2} \right).$$

Proof. To begin, we note from (3) and (36) that

$$m(e(\mu, t^k) - e(\mu, t^{k-1}), v) + \Delta t\; a(e(\mu, t^k), v; \mu) = \Delta t \left(\int_\Omega v(g(x;\mu) - g_M(x;\mu))\; b(t^k) + a_1(u_{N,M}(\mu, t^k), v; g(\cdot;\mu) - g_M(\cdot;\mu)) \right), \; \forall\, v \in W_N^u. \quad \text{(A.2)}$$

with initial condition $e(\mu, t^0) = 0$, since $u(\mu, t^0) = u_{N,M}(\mu, t^0) = 0$ by assumption. It then follows that

$$m(e(\mu, t^k), v) + \Delta t\; a(e(\mu, t^k), v; \mu) = m(e(\mu, t^{k-1}), v) + \Delta t\; a(e(\mu, t^{k-1}), v; \mu) - \Delta t\; a(e(\mu, t^{k-1}), v; \mu) + \Delta t \left(\int_\Omega v(g(x;\mu) - g_M(x;\mu))\; b(t^k) + a_1(u_{N,M}(\mu, t^k), v; g(\cdot;\mu) - g_M(\cdot;\mu)) \right), \; \forall\, v \in W_N^u. \quad \text{(A.3)}$$

Let us set now $w_N(\mu, t^k) = \pi_N[u(\mu, t^k)]$ and choose $v = e_N(\mu, t^k) \equiv w_N(\mu, t^k) - u_{N,M}(\mu, t^k)$ in (A.3). We obtain

$$\begin{aligned} & m(e_N(\mu, t^k), e_N(\mu, t^k)) + \Delta t\; a(e_N(\mu, t^k), e_N(\mu, t^k); \mu) \\ &= m(e_N(\mu, t^{k-1}), e_N(\mu, t^k)) + \Delta t\; a(e_N(\mu, t^{k-1}), e_N(\mu, t^k); \mu) - \Delta t\; a(e(\mu, t^{k-1}), e_N(\mu, t^k); \mu) \\ &\qquad + \Delta t \left(\int_\Omega v(g(x;\mu) - g_M(x;\mu))\; b(t^k) + a_1(u_{N,M}(\mu, t^k), v; g(\cdot;\mu) - g_M(\cdot;\mu)) \right) \\ &= m(e_N(\mu, t^{k-1}), e_N(\mu, t^k)) - \Delta t\; a(u(\mu, t^{k-1}) - w_N(\mu, t^{k-1}), e_N(\mu, t^k); \mu) - \Delta t\; a(e(\mu, t^{k-1}), e_N(\mu, t^k); \mu) \\ &\qquad + \Delta t \left(\int_\Omega v(g(x;\mu) - g_M(x;\mu))\; b(t^k) + a_1(u_{N,M}(\mu, t^k), v; g(\cdot;\mu) - g_M(\cdot;\mu)) \right). \end{aligned} \quad \text{(A.4)}$$

It thus follows that

$$
\begin{aligned}
& m(e_N(\mu,t^k),e_N(\mu,t^k)) - m(e_N(\mu,t^{k-1}),e_N(\mu,t^{k-1})) + \Delta t\ a(e_N(\mu,t^k),e_N(\mu,t^k);\mu) \\
& \qquad \leq \quad \Delta t\ a(u(\mu,t^{k-1}) - w_N(\mu,t^{k-1}), u(\mu,t^{k-1}) - w_N(\mu,t^{k-1});\mu) \\
& \qquad\quad +2\,\Delta t \left(\int_\Omega v(g(x;\mu) - g_M(x;\mu))\ b(t^k) + a_1(u_{N,M}(\mu,t^k), v; g(\cdot;\mu) - g_M(\cdot;\mu)) \right), \quad \text{(A.5)}
\end{aligned}
$$

which after summing from $k' = 1$ to k leads to

$$
\begin{aligned}
& m(e_N(\mu,t^k),e_N(\mu,t^k)) + \Delta t \sum_{k'=1}^{k} a(e_N(\mu,t^{k'}),e_N(\mu,t^{k'});\mu) \\
& \qquad \leq \quad \Delta t \sum_{k'=1}^{k-1} a(u(\mu,t^{k'}) - w_N(\mu,t^{k'}), u(\mu,t^{k'}) - w_N(\mu,t^{k'});\mu) \\
& \qquad\quad +2\,\Delta t \sum_{k'=1}^{k} \left(\phi_1(\mu)\,|b(t^{k'})| + \phi_2(\mu)\,\|u_{N,M}(\mu,t^{k'})\|_X \right) \varepsilon_M(\mu)\,\|e_N(\mu,t^k)\|_X, \quad \text{(A.6)}
\end{aligned}
$$

where the last inequality follows from (25) and (26). We take the square root of what we have obtained

$$
\begin{aligned}
& \left\{ m(e_N(\mu,t^k),e_N(\mu,t^k)) + \Delta t \sum_{k'=1}^{k} a(e_N(\mu,t^{k'}),e_N(\mu,t^{k'});\mu) \right\}^{1/2} \\
& \qquad \leq \quad \left\{ \Delta t \sum_{k'=1}^{k-1} a(u(\mu,t^{k'}) - w_N(\mu,t^{k'}), u(\mu,t^{k'}) - w_N(\mu,t^{k'});\mu) \right\}^{1/2} \\
& \qquad\quad + \left\{ 2\,\Delta t \sum_{k'=1}^{k} \left(\phi_1(\mu)\,|b(t^{k'})| + \phi_2(\mu)\,\|u_{N,M}(\mu,t^{k'})\|_X \right) \varepsilon_M(\mu)\,\|e_N(\mu,t^k)\|_X \right\}^{1/2}, \quad \text{(A.7)}
\end{aligned}
$$

so a triangular inequality gives

$$
\begin{aligned}
& \Bigg\{ m(u(\mu,t^k) - u_{N,M}(\mu,t^k), u(\mu,t^k) - u_{N,M}(\mu,t^k)) \\
& \qquad\qquad +\Delta t \sum_{k'=1}^{k} a(u(\mu,t^{k'}) - u_{N,M}(\mu,t^{k'}), u(\mu,t^{k'}) - u_{N,M}(\mu,t^{k'});\mu) \Bigg\}^{1/2} \\
& \qquad\quad \leq 2 \Bigg\{ m(u(\mu,t^k) - w_N(\mu,t^k), u(\mu,t^k) - w_N(\mu,t^k)) \\
& \qquad\qquad + \Delta t \sum_{k'=1}^{k} a(u(\mu,t^{k'}) - w_N(\mu,t^{k'}), u(\mu,t^{k'}) - w_N(\mu,t^{k'});\mu) \Bigg\}^{1/2} \\
& \qquad\qquad + \left\{ 2\,\Delta t \sum_{k'=1}^{k} \left(\phi_1(\mu)\,|b(t^{k'})| + \phi_2(\mu)\,\|u_{N,M}(\mu,t^{k'})\|_X \right) \varepsilon_M(\mu)\,\|e_N(\mu,t^k)\|_X \right\}^{1/2}. \quad \text{(A.8)}
\end{aligned}
$$

We now note that $\|e_N(\mu,t^k)\|_X \le \|u(\mu,t^k) - w_N(\mu,t^k)\|_X + \|e(\mu,t^k)\|_X$ and recall the identity (for $c \in \mathbb{R},\ d \in \mathbb{R},\ \varrho \in \mathbb{R}_+$)

$$2\ |c|\ |d| \le \frac{1}{\varrho^2} c^2 + \varrho^2\ d^2, \tag{A.9}$$

which we apply four times: first, with $c = \varepsilon_M(\mu)\,\phi_1(\mu)\,|b(t^k)|$, $d = \|u(\mu,t^k) - w_N(\mu,t^k)\|_X$, and $\varrho^2 = \alpha(\mu)$; second, with $c = \varepsilon_M(\mu)\,\phi_1(\mu)\,|b(t^k)|$, $d = \|e(\mu,t^k)\|_X$, and $\varrho^2 = \alpha(\mu)/8$; third, with $c = \varepsilon_M(\mu)\,\phi_2(\mu)\,\|u_{N,M}(\mu,t^k)\|_X$, $d = \|u(\mu,t^k) - w_N(\mu,t^k)\|_X$, and $\varrho^2 = \alpha(\mu)$; and fourth, with $c = \varepsilon_M(\mu)\,\phi_2(\mu)\,\|u_{N,M}(\mu,t^k)\|_X$, $d = \|e(\mu,t^k)\|_X$, and $\varrho^2 = \alpha(\mu)/8$. We can then bound the last term of (A.6) by

$$\begin{aligned}
&2\,\Delta t \sum_{k'=1}^{k} \left(\phi_1(\mu)\,|b(t^{k'})| + \phi_2(\mu)\,\|u_{N,M}(\mu,t^{k'})\|_X\right) \varepsilon_M(\mu)\,\|e_N(\mu,t^k)\|_X \\
&\qquad \le \quad \varepsilon_M(\mu)^2\,\frac{9}{\alpha(\mu)} \left(\phi_1(\mu)^2\,\Delta t \sum_{k'=1}^{k} b(t^{k'})^2 + \phi_2(\mu)^2 \Delta t \sum_{k'=1}^{k} \|u_{N,M}(\mu,t^{k'})\|_X^2\right) \\
&\qquad\quad +2\,\Delta t\,\alpha(\mu) \sum_{k'=1}^{k} \|u(\mu,t^{k'}) - w_N(\mu,t^{k'})\|_X^2 + \Delta t \frac{\alpha(\mu)}{4} \sum_{k'=1}^{k} \|e(\mu,t^{k'})\|_X^2.
\end{aligned} \tag{A.10}$$

We next use $v = u_{N,M}(\mu,t^k)$ in (36), invoke the Cauchy-Schwarz inequality for $m(u_{N,M}(\mu,t^k), u_{N,M}(\mu,t^{k-1}))$ and apply (A.9) with $c = m^{1/2}(u_{N,M}(\mu,t^k), u_{N,M}(\mu,t^k))$, $d = m^{1/2}(u_{N,M}(\mu,t^{k-1}), u_{N,M}(\mu,t^{k-1}))$, and $\varrho = 1$, to get

$$\begin{aligned}
&m(u_{N,M}(\mu,t^k), u_{N,M}(\mu,t^k)) - m(u_{N,M}(\mu,t^{k-1}), u_{N,M}(\mu,t^{k-1})) \\
&\qquad +2\,\Delta t\,a(u_{N,M}(\mu,t^k), u_{N,M}(\mu,t^k);\mu) \\
&\qquad\qquad \le \quad 2\,\Delta t \int_\Omega (u_{N,M}(\mu,t^k) g_M(x;\mu))\,b(t^k) \\
&\qquad\qquad\qquad\qquad +2\,\Delta t\,a_1(u_{N,M}(\mu,t^k), u_{N,M}(\mu,t^k); g(x;\mu) - g_M(x;\mu)) \\
&\qquad\qquad \le \quad 2\,\Delta t\,\phi_3(\mu)\,\|u_{N,M}(\mu,t^k)\|_X\,|b(t^k)| + 2\,\Delta t\,\varepsilon_M(\mu)\,\phi_2(\mu)\,\|u_{N,M}(\mu,t^k)\|_X^2 \\
&\qquad\qquad \le \quad \frac{\Delta t}{\alpha(\mu) - 2\,\phi_2(\mu)\,\varepsilon_M(\mu)} \phi_3(\mu)^2\,b(t^k)^2 + \Delta t\,\alpha(\mu)\,\|u_{N,M}(\mu,t^k)\|_X^2,
\end{aligned} \tag{A.11}$$

where the second inequality follows from (26) and (27), and the last inequality from (A.9) with $c = \phi_3(\mu)\,b(t^k)$, $d = \|u_{N,M}(\mu,t^k)\|_X$, and $\varrho = \alpha(\mu) - 2\,\phi_2(\mu)\,\varepsilon_M(\mu)$; note that $\varrho > 0$ from our assumption

on $\varepsilon_M(\mu)$. Invoking (15) and summing (A.11) from $k'=1$ to k we obtain

$$m(u_{N,M}(\mu,t^k),u_{N,M}(\mu,t^k))+\Delta t\sum_{k'=1}^{k}a(u_{N,M}(\mu,t^{k'}),u_{N,M}(\mu,t^{k'});\mu)$$
$$\leq \frac{\phi_3(\mu)^2}{\alpha(\mu)-2\,\phi_2(\mu)\,\varepsilon_M(\mu)}\,\Delta t\sum_{k'=1}^{k}b(t^k)^2. \quad \text{(A.12)}$$

From the coercivity of m and a, and our assumption on $\varepsilon_M(\mu)$ it then directly follows that

$$\begin{aligned}\Delta t\sum_{k'=1}^{k}\|u_{N,M}(\mu,t^{k'})\|_X^2 &\leq \frac{\phi_3(\mu)^2}{\alpha(\mu)(\alpha(\mu)-2\,\phi_2(\mu)\,\varepsilon_M(\mu))}\,\Delta t\sum_{k'=1}^{k}b(t^k)^2\\ &\leq \frac{2\,\phi_3(\mu)^2}{\alpha(\mu)^2}\,\Delta t\sum_{k'=1}^{k}b(t^k)^2. \quad \text{(A.13)}\end{aligned}$$

From (A.8) and invoking (A.10) and (A.13) we obtain

$$\begin{aligned}&m(e(\mu,t^k),e(\mu,t^k)+\Delta t\sum_{k'=1}^{k}a(e(\mu,t^{k'}),e(\mu,t^{k'});\mu)\\ &\leq\ 8m(u(\mu,t^k)-w_N(\mu,t^k),u(\mu,t^k)-w_N(\mu,t^k))\\ &\quad +8\Delta t\sum_{k'=1}^{k}a(u(\mu,t^{k'})-w_N(\mu,t^{k'}),u(\mu,t^{k'})-w_N(\mu,t^{k'});\mu)\\ &\quad +4\,\Delta t\,\alpha(\mu)\sum_{k'=1}^{k}\|u(\mu,t^{k'})-w_N(\mu,t^{k'})\|_X^2+\Delta t\frac{\alpha(\mu)}{2}\sum_{k'=1}^{k}\|e(\mu,t^{k'})\|_X^2\\ &\quad +\Upsilon(\mu)\,\Delta t\sum_{k'=1}^{k}b(t^{k'})^2, \quad \text{(A.14)}\end{aligned}$$

where

$$\Upsilon(\mu)=\frac{18}{\alpha(\mu)}\,\varepsilon_M(\mu)^2\left(\phi_1(\mu)^2+\phi_2(\mu)^2\frac{2\,\phi_3(\mu)^2}{\alpha(\mu)^2}\right).$$

The desired result then directly follows from the fact $w_N(\mu,t^k)$ is the projection of $u(\mu,t^k)$ with respect to the $m+\Delta t a$ norm. □